DESCRIPTION

LIGAND SCREENING APPARATUS, LIGAND SCREENING METHOD, PROGRAM AND RECORDING MEDIUM

TECHNICAL FIELD

The present invention relates to ligand screening apparatuses, ligand screening methods, programs and recording medium using protein 3D structure coordinates, and more specifically to a ligand screening apparatus, a ligand screening method, a program and a recording medium for predicting a ligand that is considered as being involved in interaction, for a protein having known 3D structure coordinate.

BACKGROUND ART

Proteins required for maintenance of biological functions such as enzymes and receptors have properties called "substrate specificity", and such proteins are classified into "Lock&Key" type wherein an active site constantly remains unchanged to details of structure of substrate molecule, and "Induced-Fit" (induced-bonding) type wherein an active site is in a random inactive state in the absence of a substrate, and the active site changes into an active state in the presence of a substrate for

capturing the coming substrate. By the term induced-fit type, such a receptor is contemplated that 3D structure of a ligand binding site changes in binding with a ligand to allow intake of the ligand.

As a computational chemical technique for screening for ligand molecules using 3D structure of protein, 3D compound database screening (Virtual Screening) as reported in AutoDocK ("Morris, G.M. Goodsell, D.S. Halliday, R.S. Huey, R. Hart, W.E. Belew, R.K. Olson, A.J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical biding free energy function. J. Comput. Chem. 19:1639-1662; Goodsell, D.S. Morris, G.M. Olson, A. J. (1996) Automated docking of flexible ligands: applications of AutoDock. J. Mol. Recognit, 9: 1-5"), DOCK ("Ewing, T.J. I Makino, S. Skillman, A. G. Kuntz I. D. (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J. Comput. Aided Mol. Des. 15: 411-28"), FlexX ("Rarey, M, Kramer, B, Lengauer, T, Klebe G. (1996) A fast flexible docking method using an incremental construction algorithm. J. Mol. Biol. 261: 470-89; Rarey, M. Wefing, S. Lengauer, T. (1996) Placement of medium-sized molecular fragments into active sites of proteins. J. Comput. Aided Mol. Des. 10: 41-54"), GOLD ("Jones, G. Willett, P. Glen, R. C. (1995) Molecular recognition of receptor sites using a genetic

algorithm with a description of desolvation. J. Mol. Biol. 245:43-53; Jones, G. Willett, P. Glen, R.C, Leach, A.R. Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol. 267: 727-48"), ADAM&EVE ("Mizutani, M.Y. Itai, A. (2004) Efficient method for high-throughput virtual screening based on flexible docking: discovery of novel acetylcholinesterase inhibitors. J. Med. Chem. 47: 4818-4828") are known. These are also called "High-performance docking study" and enable a mass-scale compound library screening. However, ability of these techniques to predict binding conformation and binding energy is poor because rough approximation is used for evaluation. In addition, since they fail to consider computational expression parameters corresponding to "induced-binding" which is very important for binding between protein and ligand, and even if such consideration is made, it is to such an extent that random numbers are generated and side chains of receptor are moved, and accuracy of computation result is not sufficient.

As a method for simulating "induced-binding" which is important for binding between protein and ligand, MD (molecular dynamic calculation), MM (molecular mechanical calculation), and MC (Monte Carlo method) are known. These methods provide relatively high accuracy, and enable

prediction of binding conformation and binding energy. Here, the technique called "molecular dynamic method (MD)" calculates dynamic structure of a molecule by sequentially solving dynamic equation based on the classical dynamics for each atom constituting the molecule, and enables simulation of dynamic behavior of a protein with high accuracy. However, it is not necessarily useful means because it requires significant time for calculation and has difficulty in handling many molecules. Further, molecular dynamic calculation executed for a target protein according to such a conventional method results in a protein 3D structure whose coordinate is largely different from those analyzed by X-ray, NMR and the like. Although such difference includes a physicochemical description of dynamic behavior of protein, it sometimes behaves contradictorily to an experimental result of dynamic behavior proved by NMR or the like, and hence it often fails to provide accurate simulation result.

As described above, in respect of the conventional "in silico screening", since computational expression parameters corresponding to "induced-binding" which is very important for binding between protein and ligand are not sufficiently considered, it does not deem that the accuracy of calculation result is adequate.

On the other hand, in molecular simulation, it is

possible to express and analyze the above induced-binding; however, significant time is required for obtaining an accurate result. Many results will be influenced by the initial structure coordinate.

Inventors of the present invention examined the way of screening for a ligand that will bind to a target protein when 3D structure of a certain protein is given. As described above, some currently available receptor-ligand binding analyzing software takes flexibility of a ligand into account, but most of such software fails to consider flexibility of a receptor. Even though there is software that considers flexibility of a receptor, such consideration just moves a side chain of the receptor by generation of random numbers, and most of software are dedicated to a Lock&Key type receptor. In such circumstances, we attempted to develop receptor-ligand binding analyzing software dedicated to an Induced-Fit type receptor.

The problem to be solved by the present invention is to provide a ligand screening apparatus, a ligand screening method, a program and a recording medium capable of screening for a ligand that binds to a certain protein which is a particularly important key to development of agricultural chemicals and pharmaceuticals and the like, with significantly higher efficiency and accuracy than

conventional methods. It is also an object to provide a ligand screening apparatus, a ligand screening method, a program and a recording medium which carry out various modifications of ligand molecule and modifications of protein such as receptor rapidly and effectively. It is also an object of the present invention to clarify a mode of interaction between ligand and protein and make the recognition mechanism of the interaction clear, thereby identifying a cause of disease, and promoting development of related drugs.

DISCLOSURE OF INVENTION

Inventors of the present invention diligently examined method of screening for a ligand that binds to a target protein when any 3D structure of such a protein is given, and finally established a ligand screening apparatus, a ligand screening method, a computer program and a recording medium as will be described later.

Here used a procedure called "molecular dynamic method (MD)" which calculates dynamic structure of a molecule by sequentially solving dynamic equation based on the classical dynamics for each atom constituting the molecule. In other words, this procedure calculates dynamic behavior based on classic dynamics in each atom constituting a certain molecule. Therefore, if one can

employ this procedure successfully, even when an induced-fit type receptor in a state that no ligand is captured is selected as an initial state, binding between the receptor and a ligand could be reconstructed. Since the MD calculation is based on the classic dynamics, it is necessary to impose certain degree of constraint to each atom. For this reason, in our developing procedure, normal mode vibration (hereinafter "normal mode") of a receptor is first analyzed to calculate fluctuation of dihedral angle of main chain of the receptor, and then MD calculation is conducted while imposing constraint on each atom based on the calculated fluctuation of dihedral angle of main chain. To be more specific, first, analysis and calculation of normal mode are conducted, and then fluctuation of dihedral angle of main chain in a steady state is calculated. Then by carrying out molecular dynamic calculation while imposing constraint on each atom based on the fluctuation, dynamic structure of the receptor is predicted more accurately. By using the dynamic structure obtained in the molecular dynamic calculation and an interaction function, receptor/ligand binding which is also applicable to an induced-fit type receptor is predicted with high accuracy. In brief, the present invention predicts receptor/ligand binding more realistically with high accuracy. Therefore, the present

invention is very useful for designing pharmaceuticals and agricultural chemicals.

To solve the objectives described above, a ligand screening apparatus according to a present invention is a ligand screening apparatus which screens for a ligand that binds to a protein when coordinate data of the protein of single chain or plural chains is given, the apparatus including: post-structural-change protein coordinate data selecting unit that effects structural change in consideration of dynamic behavior using induced-fit parameter reflecting induced fit on the coordinate data of protein and selects post-structural-change protein coordinate data; spatial point designating unit that designates a spatial point at which superposition with the ligand is to be conducted, from the post-structural-change protein coordinate data selected by the post-structural-change protein coordinate data selecting unit; interaction function calculating unit that calculates an interaction function when the protein and the ligand bind to each other using the spatial point designated by the spatial point designating unit and a ligand coordinate data of the ligand; and ligand evaluating unit that evaluates the ligand that binds to the protein based on the interaction function calculated by the interaction function calculating unit.

The ligand screening apparatus according to another aspect of the present invention is the ligand screening apparatus, wherein the interaction function calculating unit calculates the interaction function using Score (i,j) shown in Formula 1.

$$Sscore(i,j)=\sum_{ij}^{\lambda}\begin{cases}\text{when i is not equal to j} \\ \alpha\times\left[\exp\left\{-\left(d_{ij}^{s}-d_{ij}^{c}\right)^{2}\right\}-\beta\right] \Big/ \dfrac{\left(d_{ij}^{s}+d_{ij}^{c}\right)^{2}}{2} \\ \text{when i is equal to j} \\ \alpha\times(1-\beta)\end{cases} \quad \text{[Formula 1]}$$

(wherein d_{ij}^{s} is a distance between i-th spatial point and j-th spatial point in the target protein. d_{ij}^{c} is an interatomic distance between i-th atom and j-th atom in the compound. α is a coefficient for making Sscore(i,j) the maximum value when the group of spatial points in the target protein and the compound completely overlap with each other. β is a coefficient for giving a threshold value by which it can be defined as "overlapping")

The ligand screening apparatus according to another aspect of the present invention is the ligand screening apparatus, wherein the interaction function calculating unit further includes interaction function optimizing unit that carries out optimization so as to make the score of interaction function maximum.

The ligand screening apparatus according to another aspect of the present invention is the ligand screening apparatus, wherein the interaction function calculating unit further includes: interaction energy optimizing unit that calculates interaction energy with the protein for the superposed ligand after optimization of the interaction function by the interaction function optimizing unit, and optimizes the interaction energy while finely adjusting conformation of ligand 3D structure data.

The ligand screening apparatus according to another aspect of the present invention is the ligand screening apparatus, wherein the ligand evaluating unit further includes: reevaluating unit that executes the interaction function calculating unit after largely changing conformation of ligand 3D structure data following optimization by the interaction energy optimizing unit, and reevaluates the ligand that binds to the protein based on the interaction function calculated by the interaction function calculating unit.

The ligand screening apparatus according to another aspect of the present invention is the ligand screening apparatus, wherein in calculation of any one of the induced-fit parameter and the post-structural-change protein coordinate data or both, the post-structural-

change protein coordinate data selecting unit calculates normal mode for the protein coordinate data, determines intensity of fluctuation of each amino acid, and conduct molecular dynamic calculation using the intensity of fluctuation as a constraint condition.

The ligand screening apparatus according to another aspect of the present invention is the ligand screening apparatus, wherein the post-structural-change protein coordinate data selecting unit calculates a fluctuation value of dihedral angle of main chain according to normal mode calculation, and conducts molecular dynamic calculation while setting the fluctuation value as a coefficient of force K in the molecular dynamic calculation shown by Formula 2 or Formula 3.

$$\mathrm{Erot} = \mathrm{Krot}(\phi - \phi 0)^2 \qquad \text{[Formula 2]}$$

(wherein Erot represents energy of dihedral angle of main chain atom in 3D structure of a protein. ϕ represents dihedral angle of main chain atoms. $\phi 0$ represents standard value of dihedral angle of main chain atoms. Here, when a value of Krot is large, ϕ is constrained by $\phi 0$.)

$$\mathrm{Epos} = \mathrm{Kpos}(\mathrm{r} - \mathrm{r0})^2 \qquad \text{[Formula 3]}$$

(wherein Epos represents position energy of main chain atom in 3D structure of a protein. r represents coordinate of main chain atom. r0 represents standard value of coordinate of main chain atom. Here, when a value of Kpos is large, r is constrained by r0.)

The ligand screening apparatus according to another aspect of the present invention is the ligand screening apparatus, wherein the interaction function calculating unit uses the interaction function to which a dynamic property function representing dynamic property of protein is added as "elastic energy".

The ligand screening apparatus according to another aspect of the present invention is the ligand screening apparatus, wherein the interaction function calculating unit adapts "U collision" as "elastic energy" which is a function shown by Formula 4 in consideration of local flexibility of protein.

$$U_{collision} = \sum_{i=1}^{M} \sum_{j=1}^{N} \varphi(i,j)$$
$$\varphi(i,j) = K\,collision * (R\,collision(i,j) - R)^2$$

[Formula 4]

(wherein M is number of atoms in an active site that prohibit collision, N is number of atoms of ligand. When interatomic distance R between an i-th atom of a main

chain or a side chain with a little dynamic behavior in active site, and j-th atom in the ligand is not more than collision distance "Rcollision(i,j)", $\phi(i,j)$ is calculated)

The ligand screening apparatus according to another aspect of the present invention is the ligand screening apparatus, wherein the interaction function calculating unit uses the interaction function to which a normal mode analysis result or secondary structure determination result of the protein is added as a dynamic property function that represents dynamic property of protein.

A ligand screening method according to the present invention is the ligand screening method which screens for a ligand that binds to a protein when coordinate data of the protein of single chain or plural chains is given, the method including: post-structural-change protein coordinate data selecting step that effects structural change in consideration of dynamic behavior using induced-fit parameter reflecting induced fit on the coordinate data of protein and selects post-structural-change protein coordinate data; spatial point designating step that designates a spatial point at which superposition with the ligand is to be conducted, from the post-structural-change protein coordinate data selected by the post-structural-change protein coordinate data selecting step; interaction

function calculating step that calculates an interaction function when the protein and the ligand bind to each other using the spatial point designated by the spatial point designating step and a ligand coordinate data of the ligand; and ligand evaluating step that evaluates the ligand that binds to the protein based on the interaction function calculated by the interaction function calculating step.

The ligand screening method according to another aspect of the present invention is the ligand screening method, wherein the interaction function calculating step calculates the interaction function using Score (i,j) shown in Formula 1.

$$Sscore(i,j)=\sum_{ij}^{\lambda}\begin{cases}\text{when i is not equal to j} \\ \alpha\times\left[\exp\left\{-\left(d_{ij}^{s}-d_{ij}^{c}\right)^{2}\right\}-\beta\right] \Big/ \dfrac{\left(d_{ij}^{s}+d_{ij}^{c}\right)^{2}}{2} \\ \text{when i is equal to j} \\ \alpha\times(1-\beta)\end{cases} \quad \text{[Formula 1]}$$

(wherein d_{ij}^{s} is a distance between i-th spatial point and j-th spatial point in the target protein. d_{ij}^{c} is an interatomic distance between i-th atom and j-th atom in the compound. α is a coefficient for making Sscore(i,j) the maximum value when the group of spatial points in the target protein and the compound completely overlap with

each other. β is a coefficient for giving a threshold value by which it can be defined as "overlapping")

The ligand screening method according to another aspect of the present invention is the ligand screening method, wherein the interaction function calculating step further includes interaction function optimizing step that carries out optimization so as to make the score of interaction function maximum.

The ligand screening method according to another aspect of the present invention is the ligand screening method, wherein the interaction function calculating step further includes: interaction energy optimizing step that calculates interaction energy with the protein for the superposed ligand after optimization of the interaction function by the interaction function optimizing step, and optimizes the interaction energy while finely adjusting conformation of ligand 3D structure data.

The ligand screening method according to another aspect of the present invention is the ligand screening method, wherein the ligand evaluating step further includes: reevaluating step that executes the interaction function calculating step after largely changing conformation of ligand 3D structure data following optimization by the interaction energy optimizing step, and reevaluates the ligand that binds to the protein based

on the interaction function calculated by the interaction function calculating step.

The ligand screening method according to another aspect of the present invention is the ligand screening method, wherein in calculation of any one of the induced-fit parameter and the post-structural-change protein coordinate data or both, the post-structural-change protein coordinate data selecting step calculates normal mode for the protein coordinate data, determines intensity of fluctuation of each amino acid, and conduct molecular dynamic calculation using the intensity of fluctuation as a constraint condition.

The ligand screening method according to another aspect of the present invention is the ligand screening method, wherein the post-structural-change protein coordinate data selecting step calculates a fluctuation value of dihedral angle of main chain according to normal mode calculation, and conducts molecular dynamic calculation while setting the fluctuation value as a coefficient of force K in the molecular dynamic calculation shown by Formula 2 or Formula 3.

$$\mathrm{Erot} = \mathrm{Krot}(\phi - \phi 0)^2 \qquad \text{[Formula 2]}$$

(wherein Erot represents energy of dihedral angle of main chain atom in 3D structure of a protein. ϕ represents

dihedral angle of main chain atom. $\phi 0$ represents standard value of dihedral angle of main chain atom. Here, when a value of Krot is large, ϕ is constrained by $\phi 0$.)

$$Epos = Kpos(r - r0)^2 \qquad \text{[Formula 3]}$$

(wherein Epos represents position energy of main chain atom in 3D structure of a protein. r represents coordinate of main chain atom. r0 represents standard value of coordinate of main chain atom. Here, when a value of Kpos is large, r is constrained by r0.)

The ligand screening method according to another aspect of the present invention is the ligand screening method, wherein the interaction function calculating step uses the interaction function to which a dynamic property function representing dynamic property of protein is added as "elastic energy".

The ligand screening method according to another aspect of the present invention is the ligand screening method, wherein the interaction function calculating step adapts "U collision" as "elastic energy" which is a function shown by Formula 4 in consideration of local flexibility of protein.

$$U_{collision} = \sum_{i=1}^{M}\sum_{j=1}^{N}\varphi(i,j)$$

$$\varphi(i,j) = K\,collision * (R\,collision(i,j) - R)^2$$

[Formula 4]

(wherein M is number of atoms in an active site that prohibit collision, N is number of atoms of ligand. When interatomic distance R between an i-th atom of a main chain or a side chain with a little dynamic behavior in an active site, and j-th atom in the ligand is not more than collision distance "Rcollision(i,j)", ϕ(i,j) is calculated)

The ligand screening method according to another aspect of the present invention is the ligand screening method, wherein the interaction function calculating step uses the interaction function to which a normal mode analysis result or secondary structure determination result of the protein is added as a dynamic property function that represents dynamic property of protein.

A program according to the present invention is a program which makes a computer execute a ligand screening method which screens for a ligand that binds to a protein when coordinate data of the protein of single chain or plural chains is given, the method including: post-structural-change protein coordinate data selecting step that effects structural change in consideration of dynamic

behavior using induced-fit parameter reflecting induced fit on the coordinate data of protein and selects post-structural-change protein coordinate data; spatial point designating step that designates a spatial point at which superposition with the ligand is to be conducted, from the post-structural-change protein coordinate data selected by the post-structural-change protein coordinate data selecting step; interaction function calculating step that calculates an interaction function when the protein and the ligand bind to each other using the spatial point designated by the spatial point designating step and a ligand coordinate data of the ligand; and ligand evaluating step that evaluates the ligand that binds to the protein based on the interaction function calculated by the interaction function calculating step.

The program according to another aspect of the present invention is the program, wherein the interaction function calculating step calculates the interaction function using Score (i,j) shown in Formula 1.

$$Sscore(i,j)=\sum_{ij}^{\lambda}\begin{cases}\text{when i is not equal to j} \\ \alpha\times\left[\exp\left\{-\left(d_{ij}^{s}-d_{ij}^{c}\right)^{2}\right\}-\beta\right] \Big/ \dfrac{\left(d_{ij}^{s}+d_{ij}^{c}\right)^{2}}{2} \\ \text{when i is equal to j} \\ \alpha\times(1-\beta)\end{cases} \quad \text{[Formula 1]}$$

(wherein d_{ij}^{s} is a distance between i-th spatial point and j-th spatial point in the target protein. d_{ij}^{c} is an interatomic distance between i-th atom and j-th atom in the compound. α is a coefficient for making Sscore(i,j) the maximum value when the group of spatial points in the target protein and the compound completely overlap with each other. β is a coefficient for giving a threshold value by which it can be defined as "overlapping")

The program according to another aspect of the present invention is the program, wherein the interaction function calculating step further includes interaction function optimizing step that carries out optimization so as to make the score of interaction function maximum.

The program according to another aspect of the present invention is the program, wherein the interaction function calculating step further includes: interaction energy optimizing step that calculates interaction energy with the protein for the superposed ligand after optimization of the interaction function by the interaction function optimizing step, and optimizes the interaction energy while finely adjusting conformation of ligand 3D structure data.

The program according to another aspect of the present invention is the program, wherein the ligand evaluating step further includes: reevaluating step that

executes the interaction function calculating step after largely changing conformation of ligand 3D structure data following optimization by the interaction energy optimizing step, and reevaluates the ligand that binds to the protein based on the interaction function calculated by the interaction function calculating step.

The program according to another aspect of the present invention is the program, wherein in calculation of any one of the induced-fit parameter and the post-structural-change protein coordinate data or both, the post-structural-change protein coordinate data selecting step calculates normal mode for the protein coordinate data, determines intensity of fluctuation of each amino acid, and conduct molecular dynamic calculation using the intensity of fluctuation as a constraint condition.

The program according to another aspect of the present invention is the program, wherein the post-structural-change protein coordinate data selecting step calculates a fluctuation value of dihedral angle of main chain according to normal mode calculation, and conducts molecular dynamic calculation while setting the fluctuation value as a coefficient of force K in the molecular dynamic calculation shown by Formula 2 or Formula 3.

$$Erot = Krot(\phi - \phi 0)^2$$ [Formula 2]

(wherein Erot represents energy of dihedral angle of main chain atom in 3D structure of a protein. ϕ represents dihedral angle of main chain atom. $\phi 0$ represents standard value of dihedral angle of main chain atom. Here, when a value of Krot is large, ϕ is constrained by $\phi 0$.)

$$Epos = Kpos(r - r0)^2$$ [Formula 3]

(wherein Epos represents position energy of main chain atom in 3D structure of a protein. r represents coordinate of main chain atom. r0 represents standard value of coordinate of main chain atom. Here, when a value of Kpos is large, r is constrained by r0.)

The program according to another aspect of the present invention is the program, wherein the interaction function calculating step uses the interaction function to which a dynamic property function representing dynamic property of protein is added as "elastic energy".

The program according to another aspect of the present invention is the program, wherein the interaction function calculating step adapts "U collision" as "elastic energy" which is a function shown by Formula 4 in consideration of local flexibility of protein.

$$U_{collision} = \sum_{i=1}^{M}\sum_{j=1}^{N}\varphi(i,j)$$

$$\varphi(i,j) = K\,collision * (R\,collision(i,j) - R)^2$$

[Formula 4]

(wherein M is number of atoms in an active site that prohibit collision, N is number of atoms of ligand. When interatomic distance R between an i-th atom of a main chain or a side chain with a little dynamic behavior in an active site, and j-th atom in the ligand is not more than collision distance "Rcollision(i,j)", ϕ(i,j) is calculated)

The program according to another aspect of the present invention is the program, wherein the interaction function calculating step uses the interaction function to which a normal mode analysis result or secondary structure determination result of the protein is added as a dynamic property function that represents dynamic property of protein.

A computer readable recording medium according to the present invention is the computer readable recording medium in which the program according to the present invention is recorded.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 is a conceptual view showing a basic principle

of the present invention;

FIG. 2 is a view showing a dummy hydrogen atom in sp^2 orbital;

FIG. 3 is a view showing a dummy atom in a metal atom;

FIG. 4 is a view showing an initial coordinate (B) for docking a ligand into an active site based on structure-activity relationship (SAR) information;

FIG. 5 is a view showing an initial coordinate (B) for docking a ligand into an active site based on structure-activity relationship (SAR) information;

FIG. 6 is a view showing an initial coordinate (B) for docking a ligand into an active site based on structure-activity relationship (SAR) information;

FIG. 7 is a view showing an initial coordinate (B) for docking a ligand into an active site based on structure-activity relationship (SAR) information;

FIG. 8 is a view showing an initial coordinate (B) for docking a ligand into an active site based on structure-activity relationship (SAR) information;

FIG. 9 is an illustrative view of definition of angle of hydrogen bond;

FIG. 10 is an illustrative view of definition of angle in stacking;

FIG. 11 is a view showing a result of normal mode

analysis in MODEL 1 of 1LUD;

FIG. 12 is a view showing a result of comparison between parameters of MD and clustering and scores;

FIG. 13 is a view showing distribution of maximum value and minimum value of dihedral angle constraint in MD at a fixed clustering coefficient;

FIG. 14 is a view showing constraint parameter;

FIG. 15 is a view showing distribution of dihedral angle constraint parameters at a fixed clustering parameter;

FIG. 16 is a view showing distribution of dihedral angle constraint parameters at a fixed clustering parameter;

FIG. 17 is a view showing distribution of dihedral angle constraint parameters at a fixed clustering parameter;

FIG. 18 shows distribution of dihedral angle constraint parameters at a fixed clustering parameter;

FIG. 19 is a view showing a result of MD of 1 LUD (MODEL 1);

FIG. 20 is a view showing a result of comparison with a NMR structure when the MD is calculated without constraint of dihedral angle;

FIG. 21 is a view showing a result of comparison with a NMR structure when the MD is calculated with constraint

of dihedral angle;

FIG. 22 is a view showing alignment of 1CBQ;

FIG. 23 is a view showing 3D structure of 1CBQ;

FIG. 24 is a view showing 3D structure of 1CBQ;

FIG. 25 is a view showing difference between an X-ray structure and a model structure of 1CBQ by rms;

FIG. 26 is a view showing results of normal mode analysis for an X-ray structure and a model structure of 1CBQ;

FIG. 27 is a view showing results of normal mode analysis for an X-ray structure and a model structure of 1CBQ;

FIG. 28 is a view showing results of MD calculation for an X-ray structure and a model structure of 1CBQ;

FIG. 29 is a view showing alignment of 1J9G;

FIG. 30 is a view showing 3D structure of 1J9G;

FIG. 31 is a view showing 3D structure of 1J9G;

FIG. 32 is a view showing difference between an X-ray structure and a model structure of 1J9G by rms;

FIG. 33 is a view showing results of normal mode analysis for an X-ray structure and a model structure of 1J9G;

FIG. 34 is a view showing results of normal mode analysis for an X-ray structure and a model structure of 1J9G;

FIG. 35 is a view showing results of MD calculation for an X-ray structure and a model structure of 1J9G;

FIG. 36 is a view showing alignment of 1MMB;

FIG. 37 is a view showing 3D structure of 1MMB;

FIG. 38 is a view showing 3D structure of 1MMB;

FIG. 39 is a view showing difference between an X-ray structure and a model structure of 1MMB by rms;

FIG. 40 is a view showing results of normal mode analysis for an X-ray structure and a model structure of 1J9G;

FIG. 41 is a view showing results of normal mode analysis for an X-ray structure and a model structure of 1J9G;

FIG. 42 is a view showing results of MD calculation for an X-ray structure and a model structure of 1MMB;

FIG. 43 is a view showing 3D structure of dihydrofolate reductase;

FIG. 44 is a view showing a result of normal mode analysis of 1BZF (MODEL 18);

FIG. 45 is a view showing a result of MD calculation of 1BZF (MODEL 18);

FIG. 46 is a view showing structure-activity relationship information obtained from 1LUD (MODEL4);

FIG. 47 is a view showing a result of active site-ligand binding analysis in 1BZF (MODEL 18);

FIG. 48 is a view showing 1BZF (MODEL4)-ligand binding;

FIG. 49 is a view showing 1BZF (MODEL4)-ligand binding;

FIG. 50 is a view showing 1BZF (MODEL4)-ligand binding;

FIG. 51 is a view showing 3D structure of heat shock protein 90;

FIG. 52 is a view showing a result of normal mode analysis of 1YER;

FIG. 53 is a view showing a result of MD calculation of 1YER;

FIG. 54 is a view showing structure-activity relationship information obtained from 1YER;

FIG. 55 is a view showing a result of active site-ligand binding analysis in 1YER;

FIG. 56 is a view showing 1YER-ligand binding;

FIG. 57 is a view showing 1YER-ligand binding;

FIG. 58 is a view showing 3D structure of mitogen-activated protein kinase;

FIG. 59 is a view showing a result of normal mode analysis of 1A9U;

FIG. 60 is a view showing a result of MD calculation of 1A9U;

FIG. 61 is a view showing structure-activity

relationship information obtained from 1OUK;

FIG. 62 is a view showing a result of active site-ligand binding analysis in 1A9U;

FIG. 63 is a view showing 1A9U-ligand binding;

FIG. 64 is a view showing 1A9U-ligand binding;

FIG. 65 is a view showing 1A9U-ligand binding;

FIG. 66 is a view showing 3D structure of 1AIX;

FIG. 67 is a view showing a result of in silico screening;

FIG. 68 is a view showing a protein/ligand complex structure;

FIG. 69 is a view showing a protein/ligand complex structure;

FIG. 70 is a view showing a protein/ligand complex structure;

FIG. 71 is a view showing a protein/ligand complex structure;

FIG. 72 is a view showing a protein/ligand complex structure;

FIG. 73 is a view showing a protein/ligand complex structure;

FIG. 74 is a view showing a protein/ligand complex structure;

FIG. 75 is a view showing a protein/ligand complex structure;

FIG. 76 is a view showing 3D structure of SARS protease;

FIG. 77 is a view showing a result of normal mode analysis of 1UK3 (B chain);

FIG. 78 is a view showing a result of MD calculation of 1UK3 (B chain);

FIG. 79 is a view showing structure-activity relationship information obtained from 1UK4 (B chain);

FIG. 80 is a view showing a result of in silico screening in 1UK3 (B chain);

FIG. 81 is a view showing a result of comparison between 1UK3 and 1UK4;

FIG. 82 is a view showing Ranking 1 of in silico screening;

FIG. 83 is a view showing Ranking 1 of in silico screening;

FIG. 84 is a view showing structure-activity relationship information obtained from 1UK3 (B chain);

FIG. 85 is a view showing a result of comparison between 1UK3 (B chain) and 1UK4 (B chain);

FIG. 86 is a view showing a result of in silico screening executed while designating three positions in SAR;

FIG. 87 is a view showing structure-activity relationship information obtained from 1UK3 (B chain);

FIG. 88 is a view showing a result of high throughput screening executed while designating five positions in SAR;

FIG. 89 is a view showing a result of comparison between 1UK3 (B chain) and 1UK4 (B chain);

FIG. 90 is a view showing structure-activity relationship information obtained from 1UK3 (B chain);

FIG. 91 is a view showing a result of high throughput screening executed while changing the designated ligand atom type;

FIG. 92 is a view showing a result of comparison between 1UK3 (B chain) and 1UK4 (B chain);

FIG. 93 is a view showing structure-activity relationship information obtained from 1UK4 (B chain);

FIG. 94 is a view showing a result of high throughput screening executed with fixed receptor;

FIG. 95 is a view showing a result of comparison between a ligand in which 1UK3 and 1UK4 are superposed, and a ligand of screening result;

FIG. 96 is a view showing distribution of dihedral angle constraint MD parameters in 1AXJ;

FIG. 97 is a view showing a result of MD calculation of 1LUD (MODEL 1);

FIG. 98 is a view showing a result of MD calculation of 1LUD (MODEL 1);

FIG. 99 is a view showing a result of MD calculation of 1LUD (MODEL 1);

FIG. 100 is a view showing a result of MD calculation of 1LUD (MODEL 1);

FIG. 101 is a view showing a result of MD calculation of 1LUD (MODEL 1);

FIG. 102 is a view showing a result of MD calculation of 1LUD (MODEL 1);

FIG. 103 is a view showing a result of MD calculation of 1LUD (MODEL 1);

FIG. 104 is a view showing a result of MD calculation of 1LUD (MODEL 1);

FIG. 105 is a view showing a result of MD calculation of 1LUD (MODEL 1);

FIG. 106 is a view showing a result of MD calculation of 1LUD (MODEL 1);

FIG. 107 is a view showing a result of MD calculation of 1LUD (MODEL 1);

FIG. 108 is a view showing a result of MD calculation of 1LUD (MODEL 1);

FIG. 109 is a view showing a result of receptor/ligand binding in a different condition;

FIG. 110 is a view showing a result of receptor/ligand binding in a different condition;

FIG. 111 is a view showing a result of

receptor/ligand binding in a different condition;

FIG. 112 is a view showing structure-activity relationship information modified for 1BZF;

FIG. 113 is a view showing a result of ligand binding analysis of 1BZF (MODEL 18);

FIG. 114 is a view showing a result of ligand binding analysis of 1BZF (MODEL 18);

FIG. 115 is a flowchart showing one example of main process in the present system in the present embodiment;

FIG. 116 is a block diagram showing one example of configuration of the present system to which the present invention is applied;

FIG. 117 is a block diagram showing one example of configuration of an interaction function calculating unit 102c of the present system to which the present invention is applied; and

FIG. 118 is a block diagram showing one example of configuration of a ligand evaluating unit 102d of the present system to which the present invention is applied.

BEST MODE(S) FOR CARRYING OUT THE INVENTION

Exemplary embodiments of a ligand screening apparatus, a ligand screening method, a program and a recording medium of the present invention will be explained in detail with reference to the attached drawings. It is to

be noted that the present invention is not limited to these exemplary embodiments.

Several terms used herein have the following meanings unless otherwise specified.

The term "target protein" means a protein whose precise 3D structure is already determined by X-ray crystallographic analysis, NMR analysis or homology modeling method and which is an object of ligand screening.

The term "atomic coordinate" describes 3D structure in 3D space. It includes relative distances from the origin of a certain spatial point in three directions which are perpendicular to each other, and hence an atomic coordinate is described by a vector comprising three numbers per each atom existing in a protein except for hydrogen atoms.

[Basic principal of the present invention]

The basic principal of the present invention will now be explained with reference to FIG. 1. FIG. 1 is a conceptual diagram showing the basic principal of the present invention. The present invention relates to a ligand screening apparatus, a ligand screening method, a computer program and a recording medium, wherein when a protein 3D structure consisting of a single chain or plural chains is given, a parameter reflecting induced-fit from the given 3D structure of the target protein and a 3D

structure coordinate after structural change are calculated in advance by e.g., normal mode calculating method or molecular dynamic calculating method; an interaction function in binding of the target protein and other substance (ligand) is defined using the parameter and the 3D structure coordinate after structural change; and a substance (ligand) which binds to the target protein is evaluated and chosen by the interaction function.

In the present invention, first, one ligand is selected from a compound database, and 3D structure data of the ligand is acquired (Step S-1). Also 3D structure data of a target protein is acquired (Step S-2).

Then in the present invention, based on the 3D structure data of the protein, structural change considering dynamic behavior is conducted using an induced-fit parameter reflecting induced-fit, to prepare plural sets of protein coordinate data after structural change (hereinafter referred to as "post-structural-change protein coordinate data"), and one set of post-structural-change protein coordinate data is randomly chosen (Step S-3).

Next, in the present invention, a spatial point in the post-structural-change protein coordinate data to which the ligand is to be superposed is designated (Step S-4). The spatial point may be designated, for example,

by the methods (1) and (2) as will be described below.

(1) Designation of spatial point by generation of dummy atom (Step S-4-1)

Focusing on a hydrogen bond in interaction between ligand and protein, a hydrogen bond site in the protein is designated as a spatial point. The important issues in hydrogen bond are distance and angle. That is, a hydrogen bond donor (hereinafter referred to as "donor") is required for calculating an angle.

In the present invention, when there is no hydrogen atom in an active site and the ligand, a dummy hydrogen atom is generated in accordance with the following rule.

1) A dummy atom is generated in an equilateral triangle centered at sp^2 orbital atom (FIG.2). More specifically, as shown in FIG. 2, a dummy hydrogen atom (B) is generated at a free position in the equilateral triangle centered at the nitrogen atom (A) of sp^2 orbital atom.

2) As to a sp^3 orbital atom, when it is at a distance to form a hydrogen bond, it is deemed to turn so as to share the hydrogen atom, so that only the distance is considered in calculation of hydrogen bond interaction. Therefore, no dummy atom is generated for the sp^3 orbital atom.

As to metal and water, since they can mediate binding between active site and ligand binding, a dummy atom is generated at an interacting position in the manner as

described below.

1) To metal such as iron, dummy atoms are generated in a regular octahedron (FIG. 3). That is, as shown in FIG.3, dummy atoms (B) are generated at free positions in the regular octahedron centered at zinc (A).

2) To water, dummy atoms are generated in a regular tetrahedral. It is not necessary to generate a dummy atom in the direction in which it interacts with an active site.

(2) Designation of spatial point using structure-activity relationship information (Step S-4-2)

Also in the present invention, a spatial point is designated by inputting information of the following items (A) to (D) in focus of structure-activity relationship (SAR) information of ligand.

(A) Atom of active site obtained form SAR information (hereinafter referred to as "A atom"). It follows PDB format.

(B) Atom type of ligand which is expected to interact with "A atom" (hereinafter referred as "B type"). It follows MOL2 format of SYBYL.

(C) Intensity of interaction between "A atom" and "B type" (hereinafter referred to as "C intensity").

(D) Distance of interaction between "A atom" and "B type" (hereinafter referred to as "D distance") (unit: angstrom).

In the present invention, a spatial point may be

created according to the rules shown 1) to 4) below based on the input information (A) to (D) and using an initial coordinate of ligand in an active site of protein.

1) When "A atom" is donor or metal and water (when designation of SAR information on the active side end is hydrogen bond donor or metal atom), a point and a circumference at "D distance" from "A atom" with respect to the direction of dummy atoms generated in Step S-4-1 are selected as initial coordinates (FIG. 4 and FIG. 5).

2) When "A atom" is a sp^3 orbital atom (when designation of SAR information on the active site end is a sp^3 orbital atom), a circumference at "D distance" from "A atom" is selected as initial coordinate (FIG. 6).

3) When "A atom" is a hydrogen bond acceptor (hereinafter referred to as "acceptor") (when designation of SAR information on the active site end is a hydrogen bond acceptor), a position and a circumference at "D distance" on a bonding extension of "A atom" are selected as initial coordinates (FIG. 7).

4) In other cases (when designation of SAR information on the active site end is an atom other than the above), a point on the surface of a sphere centered at "A atom" and having a radius of "D distance" is selected as an initial coordinate (FIG. 8).

Here in contrast with the above 1) to 4), an initial

coordinate of ligand may be designated directly.

Returning again to FIG. 1, in the present invention, pairs of an atom in the ligand coordinate data selected in Step S-1 and a spatial point in the protein coordinate data designated in Step S-4 are randomly selected so that they are not overlapped with each other (Step S-5).

Then in the present invention, a score Sscore(i,j) which is an interaction function represented by the following formula 1 is calculated (Step S-6).

$$Sscore(i,j) = \sum_{ij}^{\lambda} \begin{cases} \text{when i is not equal to j} \\ \alpha \times \left[\exp\left\{-\left(d_{ij}^{s} - d_{ij}^{c}\right)^2\right\} - \beta\right] \Big/ \dfrac{\left(d_{ij}^{s} + d_{ij}^{c}\right)^2}{2} \\ \text{when i is equal to j} \\ \alpha \times (1-\beta) \end{cases}$$

Here, d_{ij}^{s} is a distance between i-th spatial point and j-th spatial point in target protein. d_{ij}^{C} is an interatomic distance between i-th atom and j-th atom in the compound. α is a coefficient for making Sscore(i,j) the maximum value when the group of spatial points in the target protein and the compound completely overlap with each other. β is a coefficient for giving a threshold value by which it can be defined as "overlapping".

Preferably, α is 1.5 and β is 0.8.

Next, in the present invention, adjustment (optimization) is conducted so that the score of the interaction function determined in Step S-6 is maximum (Step S-7). Here, as the procedure for maximizing the score, a simulated annealing method is exemplified. For reducing the time, it is preferred to employ a process in which Step S-5 and Step S-6 are repeated plural times (for example, 1000 times) to find a pair in which Sscore(i,j) is maximum, and a ligand is superposed to an initial coordinate based on information of the found pair.

Next, in the present invention, interaction energy with respect to the protein is calculated for the ligand which is superposed in optimization of the interaction function in Step S-7, and the resultant interaction energy is optimized while finely adjusting the conformation of the ligand coordinate data (Step S-8). The fine adjustment of ligand conformation may be conducted by translating, rotating or changing coordinates in such an extent that the angle around a single bond will not exceed 0.3 by RSMD, about the ligand coordinate calculated in Step S-7.

Here, the fine adjustment of conformation of ligand coordinate data is preferably optimized by random search, for example. In random search, minimal changes between active site of protein and ligand are repeated, for

example, 8000 times in accordance with the items 1) to 3) below, to make an optimum energy "U optimum" is minimum.

1) Up to five bonds are selected at random from rotatable bonds, and each bond is randomly rotated within the range of ±10.0 degrees for changing the conformation of ligand. This process is effected, for example every three times.

2) In each of x, y and z axial directions, ligand is randomly translated within the range of ±1.0 angstrom. This process is effected, for example every twice.

3) In each rotation center coordinate, a coordinate of the rotation center is randomly moved within the range of ±1.0 angstrom, and the ligand is randomly rotated within the range of ±5.0 degrees with respect to each of direction of three orthogonal axes. This process is effected, for example every five times.

Next, in the present invention, conformation of ligand coordinate data is largely changed, and then the process from Step S-5 is started again and the process up to Step S-8 for optimization is repeated (Step S-9). Modification of conformation may be conducted by translating, rotating or changing coordinates in such an extent that the angle around a single bond will be equal to or more than 0.3 by RSMD, about the ligand coordinate calculated in STEP S-7.

Here, optimization by largely moving conformation of

ligand is achieved by selecting five rotatable bonds at random, for example, with respect to the conformation in the "U optimized" which is energy optimized in Step S-8, and rotating them randomly in accordance with a rotation angle interval determined for each atom type. Then the process after Step S-5 is repeated, for example, 5000 times.

After changing conformation of ligand, internal energy of the ligand "U internal" is calculated, and if the value is 500.0 or more, a subsequent calculation may be skipped, and the next ligand conformation may be caused to generate. Next, in the present invention, the process from Step S-4 to Step S-9 is conducted for the plural sets of post-structural-change protein coordinate data prepared in Step S-3, and an optimum coordinate of protein-ligand complex, and optimum energy "U optimum" are calculated (Step S-10).

Next, in the present invention, the above process is conducted for every ligand in the compound database prepared in Step S-1, and a ligand which possibly binds to the target protein is selected from the compound database (Step S-11).

In the above, a basic principal of the present invention was described. In the present invention, however, when any one of a parameter reflecting induced-

fit of protein and post-structural-change 3D structure coordinate or both are calculated using molecular dynamic calculation method, normal mode with respect to 3D structure of the target protein may be calculated; fluctuations of respective amino acids may be determined; molecular dynamic calculation may be conducted while using the intensity of the fluctuation as a constraint condition; and thereby molecular dynamic calculation may be conducted so that 3D structure of protein will not largely differ from the energy optimized structure.

In the present invention, the molecular dynamic calculation according to the present molecular dynamic calculation method may be so conducted that, for example, a fluctuation value of dihedral angle of main chain atom is calculated from the normal mode calculation, and the fluctuation value is substituted into coefficient of force K in the molecular dynamic calculation as shown in Formula 2 or Formula 3.

$$Erot = Krot\,(\phi - \phi 0)^2 \qquad \text{[Formula 2]}$$

Erot represents energy of dihedral angle of main chain atom in 3D structure of a protein. ϕ represents dihedral angle of main chain atom. $\phi 0$ represents standard value of dihedral angle of main chain atom. Here, when a

value of Krot is large, ϕ is constrained by $\phi 0$.

$$Epos = Kpos\,(r - r0)^2 \qquad \text{[Formula 3]}$$

Epos represents position energy of main chain atom in 3D structure of a protein. r represents coordinate of main chain atom. r0 represents standard value of coordinate of main chain atom. Here, when a value of Kpos is large, r is constrained by r0.

In the present invention, as a target function (interaction function) in evaluation of interaction between ligand and protein, a dynamic property function that expresses dynamic properties of protein may be added to the conventional interaction energy function as "Elastic energy". As a result, it is possible to rapidly calculate interaction energy from 3D structure coordinate of protein, and to clearly describe physicochemical property regarding dynamic behavior of the protein.

Here, as the "elastic energy", the function "U collision" shown by Formula 4 below may be adapted in consideration of local flexibility of protein. In the following example, when interatomic distance R between an i-th atom of a main chain or a side chain with a little dynamic behavior in active site, and a j-th atom in a ligand is not more than collision distance

"Rcollision(i,j)", calculation of $\phi(i,j)$ is defined as follows.

$$U_{collision} = \sum_{i=1}^{M} \sum_{j=1}^{N} \varphi(i,j)$$

$$\varphi(i,j) = K\,collision * (R\,collision(i,j) - R)^2$$

M is number of atoms in active site that prohibit collision, N is number of atoms of ligand. "K collision" is preferably 1000.0. "Rcollision(i,j)" is a sum of van der Waals radii of i-th atom in the active site and j-th atom in the ligand.

Here, with respect to each atom in the active site, when weighing w(i) that allows collision is defined, the function "U collision" shown by the following Formula 5 is used. w(i) is real number ranging from 0 to 1.

$$U_{collision} = \sum_{i=1}^{M} \sum_{j=1}^{N} \varphi(i,j)$$

$$\varphi(i,j) = w(i) * K\,collision * (R\,collision(i,j) - R)^2$$

[Formula 5]

M is number of atoms in active site, N is number of atoms in the ligand. "K collision" is preferably 1000.0. "Rcollision(i,j)" is a sum of van der Waals radii of i-th atom in the active site and j-th atom in the ligand.

The "elastic energy" may be defined by using

functions shown by Formula 6 below.

$$Ev = w\,(hard\ \ shape\ \ region),$$
$$E = 0\,(soft\ \ shape\ \ region) \qquad \text{[Formula 6]}$$

Here, "hard shape region" means a part exhibiting small dynamic behavior in 3D structure of protein, and "soft shape region" means a part exhibiting large dynamic behavior. Preferably, W is a constant and 100.

In the present invention, as a function of dynamic property of protein, a result of normal mode analysis of the protein or a result of secondary structure determination may be used. In determination of secondary structure, small fluctuation for helix or sheet regions of protein, and large fluctuation for other regions are applied to the constraint condition in interaction evaluation function and molecular dynamic calculation.

Also, according to the present invention, every plural coordinates can be evaluated equally, in short time full-automatically in screening of a ligand that binds to a target protein, even when the calculated target protein (protein coordinate data) includes typical plural 3D structural coordinates, or 3D structure of the target protein consists of plural 3D structure coordinates such as analytical result of NMR spectrum.

Also, according to the present invention, (1) coordinate data of protein is subjected to structural change while dynamic behavior is considered by using an induced-fit parameter reflecting induced-fit, and post-structural-change protein coordinate data is selected; (2) from the selected post-structural-change coordinate data of protein, a spatial point at which superposition with ligand is to be executed is designated; (3) an interaction function when the protein and the ligand bind is calculated using the designated spatial point and ligand coordinate data of a ligand; and (4) a ligand which binds to the protein is evaluated based on the calculated interaction function. This is advantageous to screen for a ligand that binds to an induced-fit type receptor protein with high efficiency and accuracy while considering flexibility of receptor and flexibility of ligand.

In addition, according to the present invention, it is possible to predict a new ligand that binds to 3D structure of a target protein by acquiring a parameter reflecting dynamic behavior of the protein which is very important for binding between protein and ligand, and using a novel interaction evaluation function in relation to a ligand, which reflects dynamic behavior of the target protein. As a result, in contrast to conventional methods,

it is possible to construct 3D structures of proteins that are more reliable and suitable for design of pharmaceuticals and the like at a speed enough to keep up with enormous genomic sequences that are globally analyzed. Conventionally, in in silico screening, an algorithm capable of satisfactorily handling the induced binding that is important for interaction between protein and ligand has not been established, however, in the present invention, by employing a calculation formula which allows simple inclusion of a parameter representing "fluctuation" of protein obtainable from a normal mode calculation result or secondary structure prediction, into an interaction energy function between protein and ligand, it is possible to satisfactorily handle induced binding.

Further, in molecular dynamic simulation, the present invention is featured by conducting normal mode calculation of a target protein with regard to a parameter reflecting dynamic behavior of the target protein and to an interaction evaluation function in relation to ligand, and reflecting the calculation result into molecular dynamic calculation. Conventionally, molecular dynamic calculation is used to simulate dynamic behavior or protein, however, when molecular dynamic calculation is conducted on a target protein by a conventional method, the protein 3D structure will greatly differ from the

coordinate that is analyzed by X-ray analysis, NMR and the like. Such difference includes physicochemical description for dynamic behavior of the protein, however, the behavior is sometimes contradictory to the experimental result of dynamic behavior proved by NMR or the like, so that the simulation is not necessary accurate. For this reason, in conducting molecular dynamic calculation, it is necessary to conduct simulation while fixing 3D structure of protein to some extent, and in the present invention, we developed a measure for constraining dihedral angle of a main chain atom in energy function in molecular dynamic calculation. Further, as a constraint condition of dihedral angle, normal mode calculation of a target protein is conducted in advance for its parameter, and fluctuation of dihedral angle of main chain atom is calculated. The fluctuation is used as a parameter in such a manner that according to the intensity of the fluctuation, the constraint condition is relaxed for a region exhibiting large fluctuation, and the constraint condition is intensified for a region exhibiting small fluctuation. Therefore, according to the present invention, it is possible to describe dynamic behavior with high accuracy by conducting molecular dynamic simulation of protein under such a condition. Additionally, it is possible to acquire a coordinate

describing dynamic behavior of protein from the molecular simulation thus calculated, and by using this, it is possible to conduct ligand screening using various shapes of ligand binding sites.

As a result of the above, the present invention enables new ligands to be found that would not be found by conventional in silico screening, and enables execution of analysis of protein-ligand interaction including "induced binding" that is enabled only by time-consuming molecular simulation heretofore, in short time.

The present invention is applicable to "in silico screening" taking induced binding phenomenon into account more intensively than existent software, and achieves simplification based on correct understandings of induced binding phenomenon and hydrophobic interaction. Since the present invention is simplified, it allows handling of more target proteins by automation. As a result, it is possible to screen new possible compounds, for example, from more than a million of compound databases, and it is possible to screen possible compounds within a realistic time from a scale of databases that cannot be experimentally handled.

Further, since the present invention enables interaction analysis between protein and ligand to be conducted in a short time, interaction between many

proteins causes for example, metabolism and toxity, and drugs can be analyzed, and thus prediction of in silico metabolism and toxity of drug is enabled.

Molecules that can be handled as ligands in the present invention are understood as any substances including proteins, peptides, DNAs, drug ingredients, metals, ions, sugars, nucleic acid components and hormones because used number and kinds of ligands are not limited. The present invention enables specific molecular designing of agricultural chemicals, pharmaceuticals and the like.

In a function for evaluating interaction energy between ligand and protein, conventionally, electrostatic energy term and van der Waals term in a docking method, and an adjustment term for expressing dynamic behavior used in a soft docking method are mainly used, however, in the present invention, instead of using an adjustment term for expressing dynamic behavior used in a soft docking method or the like, a principal of elastic collision which is used in classical mechanics is applied to interaction between protein and ligand, and thus physicochemical property of interaction between protein and ligand is more clarified. This provides relationship between structural change of protein and interaction, and helps rapid and correct understanding of function of a ligand.

3D structure of protein used in the present invention

may be adapted to a 3D structure coordinate created by using an empirical modeling method (in particular, homology modeling method and threading method) of protein besides 3D structure of protein whose 3D coordinate is determined by X-ray crystalline structure analysis or the like.

[System configuration]

Now, configuration of the present system to which the present invention is applied will be explained in detail with reference to FIG. 116. FIG. 116 is a block diagram showing one example of configuration of the present system to which the present invention is applied, and only the parts in configuration that are relevant to the present invention are schematically shown.

As shown in FIG. 116, the present system is generally made up of a ligand screening apparatus 100 for evaluating and selecting a substance (ligand) that binds to a protein, and an external system 200 for providing external databases concerning ligand 3D structure data and protein 3D structure data, as well as a variety of external programs that are communicatively connected via a network 300.

The network 300 has a function of mutually connecting the ligand screening apparatus 100 and the external system

200, and is implemented by the Internet or LAN, for example.

The external system 200 is mutually connected with the ligand screening apparatus 100 via the network 300 and has a function of providing a user with external databases concerning ligand 3D structure data and protein 3D structure data as well as websites for executing various external programs. Here, the external system 200 may be implemented by a WEB server, ASP server or the like, and its hardware configuration may be implemented by commercially available workstation, personal computer and the like information processing devices and attached devices thereof. Further, each function of the external system 200 is realized by a CPU, disc device, memory device, input device, output device, communication controlling device and the like in hardware configuration of the external system 200, and programs controlling them.

The ligand screening apparatus 100 generally includes, a control unit 102 such as CPU for centrically controlling the overall ligand screening apparatus 100; a communication controlling interface unit 104 connected with a communication device (not shown) such as router connected with communication line or the like; a memory unit 106 for storing various databases and files; and an input/output controlling interface unit 108 connected with

an input device 112 and an output device 114, and these units are communicably connected via certain communication paths. Further, the ligand screening apparatus 100 is communicably connected to the network 300 via a communication device such as router and a wired or wireless communication line such as an exclusive line.

Various databases, tables and files (ligand coordinate database 106a to ligand evaluation result file 106f) stored in the memory unit 106 is a storage unit such as stationary disc device, and stores various programs, tables, files and databases, files for web pages used for various processings.

Among these constituents of the memory unit 106, the ligand coordinate database 106a is a ligand coordinate data storing unit that stores ligand coordinate data. A protein coordinate database 106b is a protein coordinate data storing unit that stores protein coordinate data. A post-structural-change protein coordinate data file 106c is a post-structural-change protein coordinate data storing unit that stores post-structural-change protein coordinate data selected by a post-structural-change protein coordinate data selecting unit 102a as will be described later. A designated spatial point file 106d is a designated spatial point storing unit that stores information concerning a spatial point designated by a

spatial point designating unit 102b as will be described later. An interaction function calculation result file 106e is an interaction function calculation result storing unit that stores information concerning calculation result of interaction function calculated by an interaction function calculating unit 102c as will be described later. A ligand evaluation result file 106f is a ligand evaluation result storing unit that stores information concerning evaluation result of ligand evaluated by a ligand evaluating unit 102d as will be described below.

The communication controlling interface unit 104 controls communication between the ligand screening apparatus 100 and the network 300 (or communication device such as router). In other words, the communication controlling interface unit 104 has a function of communicating data with other terminal via a communication line.

The input/output controlling interface unit 108 controls the input device 112 and the output device 114. Here, as the output device 114, a speaker or the like as well as a monitor (including home TV set) may be used (hereinafter, the output device 114 is sometimes referred as "monitor"). As the input device 112, a keyboard, a mouse, a microphone or the like may be used. A monitor also realizes a pointing device function together with a

mouse.

The control unit 102 has an internal memory for storing control programs such as OS (Operating System), data in need and the like, and conducts information processing for executing various processings by these programs and the like. Functionally, the control unit 102 generally includes the post-structural-change protein coordinate data selecting unit 102a, the spatial point designating unit 102b, the interaction function calculating unit 102c and the ligand evaluating unit 102d.

Among these constituents of the control unit 102, the post-structural-change protein coordinate data selecting unit 102a is a post-structural-change protein coordinate data selecting unit that calculates structural change on coordinate data of protein using an induced-fit parameter reflecting induced-fit while taking dynamic behavior into account, and selects post-structural-change protein coordinate data. The spatial point designating unit 102b is a spatial point designating unit that selects a spatial point at which superposition with ligand is to be conducted, from post-structural-change protein coordinate data selected by the post-structural-change protein coordinate data selecting unit 102a.

The interaction function calculating unit 102c is an interaction function calculating unit that calculates an

interaction function when the protein and the ligand bind using the spatial point designated by the spatial point designating unit 102b and ligand coordinate data of ligand. Here, the interaction function calculating unit 102c further includes an interaction function optimizing unit 102c1 and an interaction energy optimizing unit 102c2 as shown in FIG. 117. The interaction function optimizing unit 102c1 is an interaction function optimizing unit that optimizes so that the score of the interaction function is maximum. The interaction energy optimizing unit 102c2 is an interaction energy optimizing unit that calculates interaction energy with the protein for the superposed ligand after optimization of the interaction function by the interaction function optimizing unit 102c1, and optimizing the interaction energy while finely adjusting conformation of ligand 3D structure data.

Returning to FIG. 116, the ligand evaluating unit 102d is a ligand evaluating unit that evaluates a ligand that binds to the protein based on the interaction function calculated by the interaction function calculating unit 102c. Here, the ligand evaluating unit 102d also includes a reevaluating unit 102d1 as shown in FIG. 118. The reevaluating unit 102d1 is a reevaluating unit that reevaluates a ligand that binds the target protein based on an interaction function calculated by the

interaction function calculating unit 102c that is executed after largely changing conformation of ligand 3D structure data following optimization by the interaction energy optimizing unit 102c2.

The details of the processes executed by these units will be described later.

[System process]

Here, one example of main process of the present system in the embodiment configured as described above will be explained in detail with reference to FIG. 115, for example. FIG. 115 is a flowchart showing one example of main process of the present system in the present embodiment. Referring now to FIG. 115, ligand screening using 3D structure of protein and induced fitting will be explained.

First, the ligand screening apparatus 100 prepares database of ligand including 3-dimentional coordinate by a process of control unit 102 and stores the prepared database in the ligand coordinate database 106a of the memory unit 106 (Step S0). Here, as a database of ligand, for example, available compound databases such as ACD, imaginary compound data collected by drawing a compound and the like may be used. Preferably, database of ligand is converted into three dimensional by molecular dynamic technique.

Next, the ligand screening apparatus 100 selects 3D structure of target protein for screening the ligand database prepared in Step S0 for a specified ligand by a process of the control unit 102, acquires 3D structure data (3D coordinate) of the selected protein, and stores it in the protein coordinate database 106b of the memory unit 106 (Step S10). As the 3D coordinate, 3D structure coordinate created by PDB which is a public database or by homology modeling method is preferably used.

Next, the ligand screening apparatus 100 conducts normal mode calculation for the target protein selected in Step S10 by a process of the post-structural-change protein coordinate data selecting unit 102a, and determines fluctuation in position of main chain atom and fluctuation in dihedral angle (Step S20). More specifically, first, a parameter representing dynamic behavior of the target protein specified in Step S10 is acquired from the database of calculation result by a normal mode analysis method or a parameter is acquired by conducting secondary structure determining calculation.

First, a method of acquiring the parameter representing dynamic behavior of the protein in Step S20, by a normal mode analysis method will be explained. The normal mode analysis method is a method of approximating potential energy as a secondary function of displacement,

and solving a dynamic equation precisely, thereby analyzing microscopic vibrations around the optimized structure. The dynamic equation to be solved is the following Formula (1) or Formula (2). For details of the normal mode analysis method, see "Wilson, E. B., Decius, J. C., and Cross P. C. 1995 Molecular vibration. McGraw-Hill.".

$$\left(\sum_j T_{ij}U_{jk}\right)\omega_k^2 = \sum_j V_{ij}U_{ik} \tag{1}$$

$$TU\Lambda = VU$$
$$\Lambda_{ij} = \omega_i\delta_{ij}, U^T TU = \left(\delta_{ij}\right) \tag{2}$$

Here, ω_k is eigen value, U_{ik} is eigen vector, and δ_{ij} is delta of Kronecker. T_{ij} and V_{ij} are respectively concern motion energy E_k and potential energy V, and the following Formula (3) and Formula (4) are provided.

$$E_k = \frac{1}{2}\sum_{i,j} T_{ij}\dot{q}_i\dot{q}_j \tag{3}$$

$$U_{vdw} = K_{vdw}\sum_{i,j(>i+2)}\left\{\left(\frac{3.8}{D_{i,j}}\right)^{12} - \left(\frac{3.8}{D_{i,j}}\right)^{6}\right\} \tag{4}$$

Here, q_i is a coordinate corresponding to degree of

freedom of vibration. q_i' (it means "q_i dot" in Formula (3)) is differentiation of q_i by time. q_j is expressed by the following Formula (5).

$$q_j = q_j^0 + \sum_k A_{jk} Q_k \quad (5)$$

Here, A_{jk} is a coefficient which connects motion Q_k and individual atomic motion q_j. q_j^0 is an optimized coordinate. Q_k is normal mode shown by the following formula.

$$Q_k = \alpha_k \cos(\omega_k t + \delta_k)$$

Here, α_k and δ_k are determined as an initial condition.

Next, in Step S20, with respect to a reference protein, using the eigen value and the eigen vector obtained above, positional fluctuation of each Cα atom at a certain temperature and a certain eigen value is calculated, and this value of fluctuation is defined as a value of fluctuation of the amino acid in which the Cα is contained. As to a value of fluctuation of each amino acid in a target protein, using the alignment in Step S50, a value identical to that for the reference protein is applied as a value of fluctuation of the target protein in

a corresponding amino acid residue pair based on comparison of the target sequence and the reference sequence. For those whose values of fluctuation are not obtained, a predetermined value is applied. The value of each amino acid in the target protein thus obtained is used as a parameter representing dynamic behavior of the target protein.

Now, a method of acquiring a parameter representing dynamic behavior of protein by secondary structure determining calculation in Step S20 will be explained. Secondary structure determination is calculated from 3D structure coordinate of protein. As software, DSSP, STRIDE and the like are preferred, but basically, a method which makes determination based on angle of twist of main chain of protein and hydrogen bond pattern may be used. Here, "DSSP (Dictionary of protein secondary structure of protein)" is software that determines α-helix and β-sheet by using a PDB format file as an input file and analyzing a hydrogen pattern of main chain, an internal-rotation angle and the like. For details of the DSSP, see "Kabsch, W. & Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers, 22:2577-2637". "STRIDE (Protein secondary structure assignment from atomic coordinate)" is software that determines α-helix

and β-sheet by using a PDB format file as an input file and analyzing a hydrogen pattern of main chain, an internal-rotation angle and the like. For details of STRIDE, see "Frishman, D & Argos, P. (1995) Knowledge-based secondary structure assignment. Proteins: structure, function and genetics, 23, 566-579".

Secondary structure calculation using the above software or the like is conducted on the reference protein, and α-helix structure, β-sheet structure or loop structure formed by each amino acid is determined. As to secondary structure determination of each amino acid in a target protein, using the alignment in Step S50, the same value for the reference protein is applied as secondary structure of the target protein in a corresponding amino acid residue pair based on comparison of the target sequence and the reference sequence. For those whose secondary structures are not determined, a predetermined result is applied. The secondary structure of each amino acid in the target protein thus obtained is used as a parameter representing dynamic behavior of the target protein.

Here, in Step s20, as a parameter representing dynamic behavior of the target protein, a calculation result acquired by the normal mode analysis method of the reference protein is preferably used. As such a

calculation result, those separately stored in a database is used. Also, the secondary structure determining calculation result is used in place of normal mode analysis calculation when a reference protein for which normal mode analysis is not conducted is used.

Returning again to a main process of FIG. 115, the ligand screening apparatus 100 conducts molecular dynamic calculation using intensity of fluctuation of target protein determined in Step S20 by a process of post-structural-change protein coordinate data selecting unit 102a as a constraint condition (Step S30).

Concretely, first, positional constraint energy of main chain "U position" is introduced into Formula 6 as shown below, and minimization (condition: temperature 300K, rectangular water bath capable of placing two at least spheres of water molecule in all directions outside the surface of the receptor, force field: AMBER [S. J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, &, P. Weiner (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins J. Am. Chem. Soc. 106, 765-784]) is effected using APRICOT [Yoneda S. & Umeyama H. (1992) Free energy perturbation calculations on multiple mutation base J. Chem. Phys. 97, 6730-6736] with the variation of the initial receptor backbone constrained.

$$U\ position = K\ position * R^2 \quad (6)$$

Here, the "K position" is, for example, 300.0 and R is difference from a initial coordinate. Next, dihedral angle constraint energy "U dihedral angle" shown by Formula 7 below is introduced to APRICOT, and MD of the minimized receptor is calculated (condition: temperature 300K, rectangular water bath capable of placing two at least spheres of water molecule in all directions outside the surface of the receptor, force field: AMBER).

$$U\ dihedral\ angle = K\ dihedral\ angle * (\theta - \theta\ equivalent)^2 \quad (7)$$

θ is dihedral angle (unit: rad). For "K dihedral angle", a maximum value and a minimum value are designated so that uneven constraint that corresponds to the fluctuation of the main chain is applied to each dihedral angle within the range between the above values. Hereinafter, MD that is conducted under constraint of dihedral angle is called MD with dihedral angle constrained.

Next, dynamic structure of receptor is clustered for obtaining protein structure coordinate by MD calculation

with dihedral angle constrained. For a previously designated active site, a population made up of active sites in structures obtained by superposing receptors of every 100 femtoseconds in the course of MD and an active site in an initial structure is established. Since dynamic information of side chain is easily lost by clustering, at first, side chain dihedral angles wherein dihedral angle χ of side chain is conserved within an average angle ± 20.0 degrees in α% of the population are collected. However, when it is determined that χ closer to the main chain is not conserved, the subsequent χ is considered as not being conserved.

Next, structures that conserve every dihedral angle of side chain in collected those are extracted from the population. Then, to compare similarity of the extracted structures, when rms (root mean square) of all atoms is β angstroms or less, it is determined as the same structure, and one is deleted, and based on the finally selected structures, a receptor dynamic structure cluster is created. As to an atom constituting unconserved dihedral angle χ, it is allowed to collide with ligand in that calculation binding to active site because it is likely to vary. α and β are constants.

Returning again to the main process of FIG. 115, the ligand screening apparatus 100 designates a group of

spatial points for locating a ligand to a ligand binding site of the target protein by a process of the spatial point designating unit 102b (Step S40). Concretely, among the plural protein 3D structure coordinates created in Step S30, one is randomly selected. A spatial point in a protein coordinate is designated, for example, by the methods (1) to (4) below.

(1) Designation of spatial point by generation of dummy atom

Focusing on the hydrogen bond in interaction between ligand and protein, a hydrogen bond site in the protein is designated as a spatial point. The important issues in a hydrogen bond are distance and angle. That is, a hydrogen in hydrogen bond donor (hereinafter referred to as "donor") is required for calculating an angle.

In the present embodiment, when there is no hydrogen atom in an active site and the ligand, a dummy hydrogen atom is generated in accordance with the following rule.

1) A dummy atom is generated in an equilateral triangle centered at sp^2 orbital atom (FIG.2). More specifically, as shown in FIG. 2, a dummy hydrogen atom (B) is generated at a free position in the equilateral triangle centered at the nitrogen atom (A) of sp^2 orbital type.

2) As to a sp^3 orbital atom, when it is at a distance to form a hydrogen bond, it is deemed to rotate so as to

share the hydrogen atom, so that only the distance is considered in calculation of hydrogen bond interaction. Therefore, no dummy atom is generated for the sp^3 orbital atom.

As to metal and water, since they can mediate binding between active site and ligand binding, a dummy atom is generated at an interacting position in the manner as described below.

1) To metal such as iron, dummy atoms are generated in a regular octahedron (FIG. 3). That is, as shown in FIG.3, dummy atoms (B) are generated at free positions in the regular octahedron centered at zinc (A).

2) To water, dummy atoms are generated in a regular tetrahedral. It is not necessary to generate a dummy atom in the direction in which it interacts with an active site.

(2) Designation of spatial point using structure-activity relationship information.

Also in the present invention, a spatial point is designated by inputting information of the following items (A) to (D) in focus of structure-activity relationship (SAR) information of ligand.

(A) Atom of active site obtained from SAR information (hereinafter referred to as "A atom"). It follows PDB format.

(B) Atom type of ligand which is expected to interact with

"A atom" (hereinafter referred as "B type"). It follows MOL2 format of SYBYL.

(C) Intensity of interaction between "A atom" and "B type" (hereinafter referred to as "C intensity").

(D) Distance of interaction between "A atom" and "B type" (hereinafter referred to as "D distance") (unit: angstrom).

In the present embodiment, a spatial point may be created according to the rules shown 1) to 4) below based on the input information (A) to (D) and using an initial coordinate of ligand in an active site of protein.

1) When "A atom" is donor or metal and water (when SAR information on the active site side is designated as hydrogen bond donor or metal atom), a position and a surrounding at "D distance" from "A atom" with respect to the direction of dummy atoms generated in "(1) Designation of spatial point by generation of dummy atom" are selected as initial coordinates (FIG. 4 and FIG. 5).

2) When "A atom" is a sp^3 orbital atom (when designation of SAR information on the active site side is a sp^3 orbital atom), a surrounding at "D distance" from "A atom" is selected as initial coordinate (FIG. 6).

3) When "A atom" is a hydrogen bond acceptor (hereinafter referred to as "acceptor") (when designation of SAR information on the active site side is a hydrogen bond acceptor), a position and a surrounding at "D distance" on

a bonding extension line of "A atom" are selected as initial coordinates (FIG. 7).

4) In other cases (when designation of SAR information on the active site side is an atom other than the above), a position on the surface of a sphere with radius of "D distance" centered at "A atom" is selected as an initial coordinate (FIG. 8).

Here in contrast with the above 1) to 4), an initial coordinate of ligand may be designated directly.

Returning again to the main process of FIG. 115, the ligand screening apparatus 100 superposes each atom in the the ligand for one ligand specified in Step S0 to the group of spatial points determined in Step S30 by a process of the interaction function calculating unit 102c (Step S50). Concretely, an initial coordinate and a ligand are superposed by the following procedures (1) to (4) which are alignment creating algorithm using distance matrix (DALI) [Holm, L., & Sander, C. (1993) Protein Structure Comparison by Alignment of Distance Matrices J. Mol. Biol. 233, 123-138] modified for low molecules.

(1) It is often the case that atom types of ligand correspond to those of "B type" multiply. In view of this, a pair wherein an atom type of "B type" and that of ligand can be regarded as being identical is created by using random numbers. In such a pair, atom types of ligand

should not overlap.

(2) Since "B type" includes a plurality of initial coordinates by Step S40, selection of initial coordinate is also selected by using random numbers.

(3) Create a distance matrix of the selected initial coordinate and ligand, and calculate Sscore(i, j) which is an interaction function as follows.

$$Sscore(i,j)=\sum_{ij}^{\lambda}\begin{cases}\text{when i is not equal to j} \\ \alpha\times\left[\exp\left\{-\left(d_{ij}^{s}-d_{ij}^{c}\right)^{2}\right\}-\beta\right] \Big/ \dfrac{\left(d_{ij}^{s}+d_{ij}^{c}\right)^{2}}{2} \\ \text{when i is equal to j} \\ \alpha\times(1-\beta)\end{cases}$$

Here, d_{ij}^{s} is an distance between i-th spatial point and j-th spatial point in the target protein. d_{ij}^{s} is an interatomic distance between i-th atom and j-th atom in a compound. a is a coefficient for making Sscore(i,j) the maximum value when the group of spatial points in the target protein and the compound completely overlap with each other. β is a coefficient for giving a threshold value by which it can be defined as "overlapping".

Preferably, α is 1.5 and β is 0.8.

(4) Repeat (1) to (3) plural times (for example, 10000 times) to find a pair whose Sscore(i, j) is maximum, and superpose a ligand to an initial coordinate based on that

pair information.

Next, the ligand screening apparatus 100 acquires a parameter representing dynamic behavior of protein according to the calculation result determined in Step S20 and Step 30 for superposition in Step S50 by a process of the interaction function calculating unit 102c, and calculates interaction energy between ligand and protein using the parameter while finely adjusting conformation of the ligand (Step S60). In other words, with respect to the ligand that is superposed in Step S50, interaction energy with the protein is calculated while optimizing the conformation by fine adjustment. Fine adjustment of conformation of the ligand may be conducted by translating, rotating or changing coordinates in such a range that the angle around a single bond will not exceed 0.3 by RSMD, about the ligand coordinate calculated in Step S50.

Here, the fine adjustment of conformation of ligand coordinate data is preferably optimized by random search, for example. In random search, infinitesimal changes between active site of protein and ligand are repeated, for example, 8000 times in accordance with the items 1) to 3) below, to make an optimum energy "U optimum" be minimum.

1) Up to five bonds are selected at random from rotatable bonds, and each bond is randomly rotated within the range of ±10.0 degrees for changing the conformation of ligand.

This process is effected, for example every three times.

2) In each direction of x, y, z axes, ligand is randomly translated within the range of ±1.0 angstrom. This process is effected, for example every twice.

3) In each center coordinate of rotation, a center coordinate of the rotation is randomly translated within the range of ±1.0 angstrom, and the ligand is randomly rotated within the range of ±5.0 degrees with respect to each of direction of three orthogonal axes. This process is effected, for example every five times.

Here, optimum energy "U optimum" is defined by the following formula. The energy functions shown on the right-hand side will be individually explained below.

$$\text{U optimum} = U_{SAR} + \text{U hydrogen} + \text{U hydrophobic} + \text{U stacking} + \text{U collision} + \text{U internal}$$

As to van der Waals radius and interatomic interaction distance, references were made to AMBER99 [J. Wang, P. Cieplak & P. A. Kollam (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem, 21, 1049-1074] and MM3 parameter [Ma B., Lii J.-H., Allinger N.L. (2000) Molecular polarizabilities and induced dipole moments in molecular mechanics J. Comput. Chem. 21, 813-825].

(a) Energy function "U_{SAR}" concerning SAR information

As an index according to SAR information, energy U_{SAR} is defined as follows.

$$U_{SAR} = \sum_{i=1}^{N} \varphi(i)$$

$$\varphi(i) = K_{SAR}(i) * (R_{SAR}(i) - R)^2 - \delta$$

N is number of SAR information, R is distance from "A atom" to an interacting atom on the ligand side, K_{SAR} (i) is i-th "C intensity", R_{SAR} (i) is i-th "D distance". δ is, for example, 20.0.

(b) Energy function "U hydrogen" concerning hydrogen bond

Assuming only one hydrogen bond is formed with respect to one donor (acceptor) of ligand, an acceptor (donor) on the active site side located at shortest distance is chosen, a binding angle θ via a hydrogen (see FIG. 9, the acutest hydrogen bond angle is defined as θ when pluralhydrogen atoms are attached to a donor atom) is calculated, and in either of the two conditions described below, ϕ(i) is calculated. In FIG. 9, A is a donor, B is hydrogen, C is acceptor and θ is angle of hydrogen bond.

$$U_{hydrogen} = \sum_{i=1}^{N} \varphi(i)$$

(1) When the donor atom is sp^3 orbital type, or angle of

hydrogen bond θ is within ±30.0 degrees,

$$\text{If} R > R_{\text{hydrogen,}} \qquad \varphi(i) = -\frac{K_{\text{hydrogen}}(i)}{\left(R - R_{\text{hydrogen}}(i) + 1.0\right)}$$

$$\text{Else,} \qquad \varphi(i) = -\frac{K_{\text{hydrogen}}(i)}{\left(R_{\text{hydrogen}}(i) - R + 1.0\right)}$$

(2) When angle of hydrogen bond θ is equal to or more than ±30.0 degrees,

$$\text{If} R > R_{\text{hydrogen,}} \qquad \varphi(i) = -\frac{K_{\text{hydrogen}}(i)}{\left(R - R_{\text{hydrogen}}(i) + 1.0\right) * \theta}$$

$$\text{Else,} \qquad \varphi(i) = -\frac{K_{\text{hydrogen}}(i)}{\left(R_{\text{hydrogen}}(i) - R + 1.0\right) * \theta}$$

N is number of sum of donors and acceptors of ligand, R is distance between two atoms forming a hydrogen bond, "K hydrogen (i)" and "R hydrogen (i)" are intensity and distance of interaction of hydrogen bond determined for each atom type.

(c) Hydrophobic interaction energy function "U hydrophobic" Atoms that are capable of hydrophobic interaction in active site (side chains of ALA, CYS, PHE, ILE, LEU, MET, PRO, VAL, TRP and TYR, except of hydroxyl group of TYR) and in ligand (carbon atom) are serially

numbered, and when interatomic distance R between i-th atom in the active site and j-th atom in the ligand is within a cutoff, $\phi(i,j)$ is calculated.

$$U_{hydrophobic} = \sum_{i=1}^{M}\sum_{j=1}^{N}\varphi(i,j)$$

$$If R > R_{hydrophobic}(i,j), \varphi(i,j) = -\frac{K_{hydrophobic}(i,j)}{\left(R - R_{hydrophobic}(i,j) + 1.0\right)}$$

$$Else, \quad \varphi(i,j) = -K_{hydrophobic}(i,j)$$

M is number of atoms in active site that are capable of hydrophobic interaction, N is number of atoms in ligand site that are capable of hydrophobic interaction. "K hydrophobic (i,j)" and "R hydrophobic (i,j)" are intensity and distance of hydrophobic interaction determined for each atom type. The cutoff is, for example, 8.0 angstroms.

(d) Stacking energy function "U stacking"

Atoms forming an aromatic ring in active site and ligand are serially numbered, and a center coordinate of aromatic ring is calculated for active site. When interatomic distance R between i-th atom in active site and j-th atom in ligand is within a cutoff, taking a center coordinate of aromatic ring formed by the i-th atom as i', and taking an atom in ligand which is closest from j-th atom and belonging to the same aromatic ring as j', $\angle ii'j=\theta_{i'j}$, $\angle i'ij=\theta_{ij}$, $\angle ii'j'=\theta_{i'j'}$, $\angle i'ij'=\theta_{ij'}$ are

calculated (FIG. 10), and when $\theta_{i'j}$ and θ_{ij} are 90.0 degrees ±10.0 degrees, "R boundary" and "θ boundary" are determined, and in either of the two conditions described below, $\phi'(i,j)$ is calculated. In FIG. 10, i' represents center of aromatic ring in active site, i represents aromatic ring atom in active site, j and j' represent aromatic ring atoms in ligand.

$$U_{stacking} = \sum_{i=1}^{M}\sum_{j=1}^{N}\varphi(i,j)$$

If $R\,boundary < 0.0$,

$$\varphi(i,j) = -K\,stacking(i,j) * R\,boundary$$

Else

$$\varphi(i,j) = -K\,stacking(i,j) * \theta\,boundary$$

$$R\,boundary = 1.0 - (R\,stacking(i,j) - R)^2$$

$$\theta\,boundary = |1.0 - \Theta|$$

$$\Theta = \frac{\pi}{180.0} * (\theta - 90.0)^2$$

M is number of atoms forming an aromatic ring in active site, N is number of atoms forming aromatic ring in ligand, and "K stacking (i,j)" and "R stacking (i,j)" are distance and intensity of stacking determined for each atom type. π is the ratio of the circumference of a circle to its diameter, and θ is an angle at which Θ is minimum in $\theta_{i'j'}$ and $\theta_{ij'}$. The cutoff is, for example, 5.0 angstroms.

(e) Intermolecular collision energy function (elastic collision energy function) "U collision"

As "elastic energy", the following function "U collision" may be applied while taking local flexibility of protein into account. When interatomic distance R between an i-th atom of a main chain or a (conserved) side chain atom with a little dynamic behavior in an active site, and j-th atom in the ligand is not more than collision distance "Rcollision(i,j)", calculation of $\phi(i,j)$ is defined as follows.

$$U_{collision} = \sum_{i=1}^{M}\sum_{j=1}^{N}\varphi(i,j)$$

$$\varphi(i,j) = K\,collision * (R\,collision(i,j) - R)^2$$

M is number of atoms in an active site that prohibit collision, N is number of atoms of ligand. "K collision" is preferably 1000.0. "Rcollision(i,j)" is a sum of van der Waals radii of i-th atom in the active site and j-th atom in the ligand.

Here, with respect to each atom in the active site, when weighting w(i) that allows collision is defined, the function "U collision" shown by the following Formula is used. w(i) is real number ranging from 0 to 1.

$$U_{collision} = \sum_{i=1}^{M}\sum_{j=1}^{N}\varphi(i,j)$$
$$\varphi(i,j) = w(i) * K\,collision * (R\,collision(i,j) - R)^2$$

M is number of atoms in an active site. N is number of atoms in ligand. "K collision" is preferably 1000.0. "Rcollision(i,j)" is a sum of van der Waals radii of i-th atom in active site and j-th atom in ligand.

(f) Ligand internal energy "U internal"

Since there is a case that a bond is broken due to errors that occur when a rotatable bond is changed little by little, calculations for "ϕ bond length (i)" and "ϕ collision (i,j)" are defined so as to prevent occurrence of atomic collision within a ligand.

$$U_{internal} = \sum_{i=1}^{L}\varphi_{bond\,length}(i,j) + \sum_{i=1}^{M}\sum_{j=1}^{N}\varphi_{collision}(i,j)$$
$$\varphi\,bond\,length(i) = K\,bond\,length * \{1000.0 * (R\,bond\,length(i) - R_1)\}^2$$
$$\varphi\,collision(i,j) = K\,collision * (R\,collision - R_2)^2$$

L is number of rotatable bonds, M is number of atoms of ligand, and N is number of non-binding atoms of i-th atom. Preferably, "K bond length" is 100.0. "R bond

length (i)" is bond length of initial structure. Preferably, "K collision" is 150.0 and "R collision" is 2.2 angstroms. R_1 and R_2 are distances between two atoms.

Returning again to the description of the main process of FIG. 115, by a process of the interaction function calculating unit 102c, the ligand screening apparatus 100 largely changes conformation of ligand coordinate data with respect to the ligand determined in Step S50, restarts the flow from Step S50, and repeats the process from Step S60 to Step S70 (optimization). Change of conformation may be conducted by translating, rotating or changing coordinates in such a range that the angle around a single bond will be equal to or more than 0.3 by RSMD, about the ligand coordinate calculated in Step S50.

Here, optimization by largely changing conformation of ligand is achieved by selecting five rotatable bonds at random, for example, with respect to the conformation in the "U optimized" which is energy optimized in Step 50, and rotating them randomly in accordance with a rotation angle interval determined for each atom type. Then the process subsequent to Step S50 and Step S60 is repeated, for example, 5000 times.

After changing conformation of ligand, internal energy of the ligand "U internal" is calculated, and if the value is 500.0 or more, a subsequent calculation may

be skipped, and the next ligand conformation may be caused to generate.

Next, the ligand screening apparatus 100 determines interaction energy between target protein and ligand that is obtained up to Step S70 by a process of interaction function calculating unit 102c (Step S80). Concretely, a complex coordinate between protein and ligand which provide optimum value in "U optimum" from Step S40 to Step S70, as well as optimum energy "U optimum" are calculated.

Next, the ligand screening apparatus 100 returns to Step S40 by a process of the control unit 102, selects another ligand in Step S0, and conducts the calculations up to Step S80 by processes of the respective processing units (Step S90). Here, the processes from Steps S40 to Step S90 are conducted for all ligands in the compound database in Step S0.

Next, the ligand screening apparatus 100 compares the interaction energies determined in Step S90 for the ligands in Step S0 by a process of ligand evaluating unit 102d, and selects a ligand that is expected to bind the target protein (Step S100). Concretely, based on a complex coordinate between protein and ligand and optimum energy "U optimum" evaluated up to Step S90, a compound (ligand) which has a possibility to bind to the protein is selected from the database in Step S0.

Now we end description of the main process of the system.

As described above, according to the ligand screening apparatus 100, it is possible to evaluate and select a substance (ligand) that binds to a target protein by means of interaction function. Concretely, 3D structure of the target protein is subjected to normal mode calculation, intensity of fluctuation of each amino acid is determined, and molecular dynamic calculation is conducted using the intensity of fluctuation as a constraint condition. Accordingly it is possible to conduct molecular dynamic calculation while preventing the 3D structure of the protein from largely deviating from the structure at which energy is minimized. It is also possible to clearly describe the physicochemical property concerning dynamic behavior of a protein. Further, it is possible to use a result of normal mode analysis or a result of secondary structure determination of protein as a function expressing dynamic property of protein. Also when there exist plural 3D structural coordinates as is the case for analytical result of NMR spectrum, it is possible to evaluate all of the plural coordinates equally, in short time and full-automatically for screening for a ligand that binds to the target protein.

Further, according to the ligand screening apparatus

100, when arbitrary 3D structure of a protein comprising an any not only single but also plural chain(s) is given, a parameter reflecting induced-fit from the 3D structure of the protein and post-structural-change 3D structure coordinate are calculated in advance by a normal mode calculation method or by a molecular dynamic calculation method; an interaction function when the protein binds to other substance is defined using the above parameter and the post-structural-change 3D structure coordinate; and a substance that binds to the protein is evaluated and selected based on the interaction function by means of computer program.

Further, according to the ligand screening apparatus 100, when a ligand that binds to the protein is selected, the series of processes shown in (0) to (8) are executed full-automatically or manually.

(0) Select one ligand from compound database. As 3D structure of target protein, a plurality of structural-change coordinates considering dynamic behavior using a parameter reflecting induced-fit are prepared, and one of them is selected at random.

(1) Designate a spatial point in the protein at which superposition is to be conducted.

(2) Select a pair of an atom in the ligand selected in (0) and a spatial point designated at random in (1) so that no

overlapping occurs.

(3) Calculate the following Sscore(i,j).

$$Sscore(i,j)=\sum_{ij}^{\lambda}\begin{cases}\text{when i is not equal to j} \\ \alpha\times\left[\exp\left\{-\left(d_{ij}^{s}-d_{ij}^{c}\right)^{2}\right\}-\beta\right] \Big/ \dfrac{\left(d_{ij}^{s}+d_{ij}^{c}\right)^{2}}{2} \\ \text{when i is equal to j} \\ \alpha\times(1-\beta)\end{cases}$$

d_{ij}^{S} is distance between ith spatial point and j-th spatial point in the protein. d_{ij}^{C} is interatomic distance between i-th atom and j-th atom in the compound. α is a coefficient for making Sscore(i,j) the maximum value when the group of spatial points in the target protein and the compound completely overlap with each other. β is a coefficient for giving a threshold value by which it can be defined as "overlapping".

(4) Make adjustment so that the score in (3) is maximum.

(5) Optimize interaction energy with the protein for the ligand superposed in (4) by fine adjustment of conformation.

(6) Largely change the conformation of the ligand, and restart from (2) and repeat the process up to (5) to achieve optimization.

(7) Conduct the course of (1) to (6) on the plurality of

structural-change coordinates prepared in (0), and calculate an optimum protein-ligand complex coordinate and optimum energy "U optimum".

(8) Conduct the course of (1) to (7) on every ligand in the compound database prepared in (0), and select a ligand that has a possibility to bind to the protein from the compound database.

Also according to the ligand screening apparatus 100, when a parameter reflecting induced-fit of protein and a post-structural-change 3D structure coordinate are calculated by using a molecular dynamic calculation method, 3D structure of the target protein is subjected to normal mode calculation, intensity of fluctuation of each amino acid is determined, and molecular dynamic calculation is conducted using the intensity of fluctuation as a constraint condition. This realizes molecular dynamic calculation without causing large difference in 3D structure of the protein from the structure minimized by energy optimization "U hydrogen".

Also according to the ligand screening apparatus 100, as a target function for evaluation of interaction between ligand and protein, a function that expresses dynamic characteristic of protein is added as elastic energy to a conventional interaction energy function, and interaction energy is rapidly calculated from the 3D structure

coordinate of protein, and physicochemical property concerning dynamic behavior of protein is clearly described.

Also according to the ligand screening apparatus 100, as a function expressing dynamic characteristics of protein, a result of normal mode analysis or a result of secondary structure determination of protein is used.

Also according to the ligand screening apparatus 100, when the calculated protein has typical plural 3D structure coordinates, or plural 3D structural coordinates as is the case for analytical result of NMR spectrum, it is possible to evaluate all of the plural coordinates equally, in short time and full-automatically for screening for a ligand that binds to the target protein.

[First example]

(Determination of parameter coefficient in MD with dihedral angle constrained and clustering)

Using the ligand screening apparatus 100 according to the above embodiment, a fluctuation value of dihedral angle was calculated by a normal mode analysis. As shown in this first example, a fluctuation value of dihedral angle was adapted as a constraint condition in molecular dynamic calculation as "K dihedral angle" in the following formula.

$$U\ dihedral\ angle = K\ dihedral\ angle * (\theta - \theta\ equivalent)^2$$

θ is dihedral angle (unit: rad). In practice, uneven constraint was applied to each dihedral angle such that it corresponds to fluctuation of dihedral angle of main chain within a range between a maximum value and a minimum value of "K dihedral angle" by designating such values. Accordingly, first example aims at determining appropriate maximum value and minimum value for "K dihedral angle".

Further, after molecular dynamic calculation, post-structural-change coordinates were subjected to cluster analysis, and a representative structure was selected. At this time, for an active site that is designated in advance, active sites of structures obtainable by superposing receptors of every 100 femtoseconds during MD, and an active site of initial structure were considered as a population. Concretely, since dynamic information of side chain is easily lost by clustering, at first, side chain dihedral angles wherein dihedral angle χ of side chain is conserved within an average angle ± 20.0 degrees in α% of the population are collected. However, when it is determined that χ closer to the main chain is not conserved, the subsequent χ is considered as not being conserved. Next, structures that satisfy the condition

into every side chain dihedral angle (conserved side chain dihedral angle) in collected structure were extracted from the population. Then, to compare similarity of the extracted structures, when the rms (root mean square) of all atoms was β angstroms or less, it was regarded as the same structure, and one was deleted, and based on the finally selected structures, a receptor dynamic structure cluster was created. As to an atom forming unconserved dihedral angle χ, it was allowed to collide in the active site and ligand binding calculation because it was easy to change. α and β are constants, and first example also aims at determining appropriate α and β.

First example also aims at obtaining best dynamic structure of main chain in an active site being in contact with a ligand. For this purpose, in calculation of rms (root mean square), only four main chain atoms (N, Cα, C, O) in the active site were considered as objectives.

Supposing that as a maximum and a minimum values of K dihedral angle and clustering coefficients α and β, those capable of reproducing the structure analyzed by NMR are appropriate, first, we conducted normal mode analysis using MODEL 1 of dihydrofolate reductase (DHFR, PDB code: 1LUD) whose structure was analyzed to determine a fluctuation value by NMR, and then conducted molecular dynamic calculation. After the molecular dynamic

calculation, we conducted clustering of receptor dynamic structure. Receptor residues located within 6 angstroms from each atom in the ligand contained in 1LUD (MODEL 1) were defined to form an active site. Further, for MD, results of 0 to 0.1 nanosecond. were used. Here, MD was conducted exhaustively for constraint ranging from a minimum to a maximum value of 0 to 1000 (intervals of 100), clustering coefficient α ranging from 0% to 90% (intervals of 10%), coefficient β ranging from 0.1 angstrom to 0.6 angstrom (intervals of 0.1 angstrom) with reference to the NMR structure average rms, and coefficients were determined by comparing every result with the NMR structure in 1LUD.

As a reference for determining coefficient β in receptor dynamic structure clustering, NMR structure average value was determined. In NMR structures in PDB (the Protein Data Bank), receptor is a simple protein and in one PDB file more than ten patterns of NMR structures were found . We tried to determine a NMR structure average rms of active site for 117 kinds of substrates including a ligand.

First, in MODEL 1, receptor residues contained within 6 angstroms radius from each atom in the the ligand were defined as forming an active site. Then for each of structure other than MODEL 1, rms with respect to the

active site of MODEL 1 was obtained, and then average rms thereof was determined. When the average rms is equal to or more than 1.0 angstrom, since it can be considered as an apparent dynamic structure, such a PDB file was excluded from the objective. As a result, objective PDB files were reduced to 71 kinds. Average rms of 71 kinds were further averaged to give NMR structure average rms. The NMR structure average rms obtained in this manner was 0.62.

As to determination of appropriate maximum and minimum values for "K dihedral angle" and determination of coefficients α and β in clustering, each parameter value was compared with NMR structure.

Since 1LUD includes 24 kinds of MODEL, and first example uses MODEL 1 as an objective, active sites of 23 kinds other than MODEL 1 were considered as true structures (observed experimentally). In each receptor dynamic structure cluster outputted as a result of calculation, rms between MODEL 1 and each true structure was calculated. The minimum rms among the calculated rms was set as "RMS minimum", and an average value of "RMS minimum" obtained from each receptor dynamic structure cluster was set as a score. Then a parameter at which the score is minimum was adopted.

In FIG. 11, a result of normal mode analysis in MODEL

1 of 1LUD is shown, and comparison results of score and parameter are shown in FIGs. 12, 13, 15-18. In FIG. 11, intensity of fluctuation in dihedral angle ϕ (orange A), ψ (green B) is shown. The closer to 0.0 is intensity of the fluctuation, the stronger the constraint of dihedral angle becomes in the molecular dynamic (MD) calculation. Secondary structures determined by STRIDE are also shown: α-helix (red D) and β-sheet (blue D). The purple C shows an active site. In FIG. 12, the normal mode analysis demonstrated that coefficient α is preferably 70%. However, since there are cases that clustering accuracy was deteriorated at 70% when generalization was applied, in first example, 80% was selected for a value of coefficient α. As shown in FIG. 13 and FIGs. 15-18, clustering coefficients α and β were fixed respectively at 80.0% and 0.4 angstrom. The score decreases as it comes blacker.

These results demonstrated that the values of FIG. 14 are optimum as a constraint condition for reducing the score. As to validation of these values, Cα atom, side chain and all atoms besides the main chain atom were examined and the parameter values in FIG. 14 are proved to be optimum.

[Second example]

(Difference in molecular dynamic calculation in the presence/absence of constraint parameters)

Molecular dynamic calculation adopting the constraint parameters calculated by the aforementioned ligand screening apparatus 100 was conducted until 2.0 nanoseconds. Then, how the structure changed in comparison with the case where the constraint parameters were not adopted in dynamic behavior of main chain atom in the active site was examined.

Case 1)

Examination was made on dihydrofolate reductase (MODEL 1 of 1LUD). Examination results are shown in FIGs. 19 to 21. As the result of normal mode calculation, values that were determined in the above first example were adapted.

In FIG. 19, with respect to MODEL 1 of each of 24 kinds of model structures described in PDB file of 1LUD, rms of main chain atom in active site was calculated, and the average rms was displayed by dotted lines. In the case where dihedral angle is constrained (A), and in the case where dihedral angle is not constrained (B), difference of main chain atom in active site from its initial structure was represented by rms.

FIG. 20 shows a comparison result with NMR structure when MD is calculated without constraint of dihedral angle.

In FIG. 20, NMR structure (1lud) is displayed in white, MD structure (1lud) id displayed in black.

In Table 1, a comparison result with NMR structure when MD is calculated without constraint of dihedral angle is shown.

(Table 1)

	ACTIVE SITE	ENTIRETY
Cα ONLY	3.8903	0.2919
MAIN CHAIN	3.8642	0.3335
ENTIRETY	4.447	0.1398RMS

In FIG. 21, a comparison result with NMR structure when MD is calculated with constraint of dihedral angle is shown.

In Table 2, a comparison result with NMR structure when MD is calculated with constraint of dihedral angle is shown.

(Table 2)

	ACTIVE SITE	ENTIRETY
Cα ONLY	0.6398	0.1194
MAIN CHAIN	0.6933	0.1053
ENTIRETY	1.2379	0.2157RMS

Case 2)

Here, we verified dependency on initial structure and presence/absence of constraint while selecting a structure (model structure) obtained by modeling according to FAMS [Ogata K., Umeyama H. (2000) An automatic homology modeling method consisting of database searches and

simulated annealing J. Mol. Graphics Mod. 18, 258-272], and X-ray structure as initial structures. Receptor residues contained within 10 angstroms radius from each atom in the ligand were defined as forming an active site.

X-ray structure (3D structure) of cellular retinoic acid binding protein type II(CRABP-II) (PDB code:1CBQ) was used. As a reference protein, intestinal fatty acid binding protein (PDB code:1ICM) exhibiting 32.1% homology was selected, and a model structure was constructed by alignment of FIG. 22. In FIGs. 23, 24, and 25, results of X-ray structure comparison with model structure are shown.

In FIG. 23, 3D structure (X-ray structure (red A) and model structure (blue B)) of 1CBQ are shown. In FIG. 24, structure of 6-(2,3,4,5,6,7-hexahydro-2,4,4-trimethyl-1-methyleneinden-2-yl)-3-methylhexa-2,4-dienoic acid which is a substance shown in green in FIG. 23 is shown. In FIG. 25, difference between X-ray structure and model structure of 1CBQ is shown by rms.

FIG. 26 is a view showing a result of normal mode analysis of X-ray structure of 1CBQ, and FIG. 27 is a view showing a result of normal mode analysis of model structure of 1CBQ. In FIG. 26 and FIG. 27, intensity of fluctuation in dihedral angle ϕ (orange A), ψ (green B) is shown. The closer to 0.0 is intensity of the fluctuation, the stronger the constraint of dihedral angle becomes in

the molecular dynamic (MD) calculation. Secondary structures determined by STRIDE are also shown: α-helix (red D) and β-sheet (blue D).

In FIG. 28, results of molecular dynamic (MD) calculation of X-ray structure and model structure of 1 CBQ are shown. Rms of a main chain atom in active site between X-ray structure and model structure was determined. As shown in FIG. 28, A: initial structure is X-ray structure, without constraint of dihedral angle; B: initial structure is X-ray structure, with constraint of dihedral angle; C: initial structure is model structure, without constraint of dihedral angle; and D: initial structure is model structure, with constraint of dihedral angle.

Case 3)

X-ray structure of Flavodoxin (PDB code:1J9G) was used. As a reference protein, flavodoxin (PDB code: 1AHN) exhibiting 29.2% homology was selected, and model structure was constructed by alignment of FIG. 29. In FIG. 29, alignments of 1J9G and 1AHN are shown. In FIG. 30, 3D structures (X-ray structure (red A) and model structure (blue B)) of 1J9G are shown. In FIG. 31, structure of flavin mononucleotide which is a substance viewed in green C is shown in FIG. 30. In FIG. 32, difference between X-

ray structure and model structure of 1J9G is shown by rms.

In FIG. 33, a result of normal mode analysis of X-ray structure of 1J9G, and in FIG. 34, a result of normal mode analysis of model structure of 1J9G are shown. In FIG. 33 and FIG. 34, intensity of fluctuation in dihedral angle ϕ (orange A), ψ (green B) is shown. The closer to 0.0 is intensity the fluctuation, the stronger the constraint of dihedral angle becomes in the molecular dynamic (MD) calculation. Secondary structures determined by STRIDE[8] are also shown: α-helix (red D) and β-sheet (blue D). The purple C shows an active site.

In FIG. 35, results of molecular dynamic (MD) calculation of X-ray structure and model structure of 1J9G are shown. Rms of a main chain atom in active site between X-ray structure and model structure was determined. In FIG. 35, A: initial structure is X-ray structure, without constraint of dihedral angle; B: initial structure is X-ray structure, with constraint of dihedral angle; C: initial structure is model structure, without constraint of dihedral angle; and D: initial structure is model structure, with constraint of dihedral angle.

Case 4)

X-ray structure of Matrix metalloproteinase-8(MMP-8) (PDB code: 1MMB) was used. As a reference protein, MMP-3

(PDB code: 1B3D) exhibiting 55.0% homology was selected, and a model structure was constructed by alignment of FIG. 36. In FIG. 36, alignments of 1MMB and 1B3D_A are shown.

In FIG. 37, 3D structures (X-ray structure (red A) and model structure (blue B)) of 1MMB are shown. In FIG. 38, structure of batimastat which is a substance viewed in green C is shown. In FIG. 39, difference between X-ray structure and model structure of 1MMB is shown by rms.

In FIG. 40, a result of normal mode analysis of X-ray structure of 1MMB, and in FIG. 41, a result of normal mode analysis of model structure of 1MMB are shown. As shown in FIG. 40 and FIG. 41, intensity of fluctuation in dihedral angle ϕ (orange A), ψ (green B) is shown. The closer to 0.0 is intensity of the fluctuation, the stronger the constraint of dihedral angle becomes in the molecular dynamic (MD) calculation. Secondary structures determined by STRIDE[8] are also shown: α-helix (red D) and β-sheet (blue D). The purple C shows an active site.

In FIG. 42, results of molecular dynamic (MD) calculation of X-ray structure and model structure of 1MMB are shown. Rms of a main chain atom in active site X-ray structure and model structure was determined. As shown in FIG. 42, A: initial structure is X-ray structure, without constraint of dihedral angle; B: initial structure is X-ray structure, with constraint of dihedral angle; C:

initial structure is model structure, without constraint of dihedral angle; and D: initial structure is model structure, with constraint of dihedral angle.

As shown in Cases 1) to 4), the result of molecular dynamic calculation adapting constraint parameters exhibit less significant structural change compared to the case where constraint parameters are not adapted. This reveals that significant structural change can be reasonably constrained and ideal structure coordinates can be obtained by adopting constraint parameters in molecular dynamic method which results in large structural change due to theory of classical mechanics. When homology is high, the accuracy of modeled structure by FMAS is also improved. That is, since structure similar to X-ray structure is obtained, the present invention may be applied for mutation proteins substituted by several amino acids.

[Third example]

(Verification of protein/ligand complex model)

By means of the ligand screening apparatus 100 in the above described embodiment, 3D structure of a ligand complex that binds to the target protein was predicted. In third example, prediction accuracy of the 3D structure coordinate of complex was examined. For this verification,

an induced-fit type protein was used, which is known 3D structure of complex and has variable conformation of active site which varies depending on presence/absence of ligand or type of ligand. Residues located within 10 angstrom radius from each atom in the ligand were defined to form an active site of protein. Since it turned out that the structure is kept substantially equivalent in MD which uses X-ray structure or NMR structure as initial structure, we decided to conduct MD until 1.0 nanosecond. In the calculation, however, hydrogen atoms were excluded. Construction of complex model was conducted in accordance with the above embodiment.

Case 1)

1BZF and 1LUD, dihydrofolate reductase (DHFR) have 100% homologous but are different conformations of active site because separate ligands bind to protein. 1BZF (MODEL 18) was selected as an initial structure, and using 2,4-diamino-5-(3, 4, 5-trimethoxy-benzyl)-pyrimidin-1-ium (FIG. 49), a protein/ligand complex model was created by the ligand screening apparatus 100 in the embodiment as described above, and examination was made by comparison with 1LUD (MODEL4) which is the true structure (FIG. 43). In FIG. 43, 3D structure of dihydrofolate reductase was shown. In FIG. 43, receptor (green A) and ligand (red B)

of 1LUD (MODEL 4), as well as receptor (blue C) and ligand (light blue D) of 1BZF (MODEL 18) were shown.

In FIG. 44, analysis results of normal mode calculation of 1BZF was shown. In FIG. 44, intensity of fluctuation in dihedral angle ϕ (orange A), ψ (green B) is shown. The closer to 0.0 is intensity of the fluctuation, the stronger the constraint of dihedral angle becomes in the molecular dynamic (MD) calculation. Secondary structures determined by STRIDE are also shown: α-helix (red D) and β-sheet (blue D). The purple C shows an active site.

In FIG. 45 and FIG. 47, results of molecular dynamics with dihedral angle constrained of 1BZF were shown. In FIG. 45, rms with true structure, 1LUD (MODEL4) in the active site is shown. In FIG. 45, A is of main chain atom, B is of side chain atom, and C is of whole atom. FIG. 47 is a result of analysis of ligand binding to active site in 1BZF (MODEL 18). These are results of binding analyses when MD calculation was effected until 0.1 or 1.0 nanosecond, and dynamic structures of receptor are extracted at interval of 100 femtoseconds in the former and 100 and 1000 femtoseconds in the latter, three separate populations are created by clustering those. Evaluation was made based on rms with true structure of ligand in active site. In FIG. 46, parameter values for

designating spatial points in ligand docking are shown. FIG. 46 shows information of structure-activity relationship obtained from 1LUD (MODEL4).

In FIGs. 48 to 50, comparisons between protein/ligand complex and true structure are shown. FIG. 48 shows binding between active site and ligand using a population created by clustering dynamic structure of receptor after those were extracted at interval of 100 femtoseconds within 0-1.0 nanosecond in the MD. Green A shows 1LUD (MODEL4) in true structure, blue B shows 1BZF (MODEL 18) in initial structure, and red C shows optimum structure in ligand binding, "by element" color D shows ligand in true structure; and light blue E shows ligand by calculation result. Rms of ligand was 0.9614. By causing the ligand shown in FIG. 50 to bind, induction of 0.2791 by rms occurred in the main chain atom in the active site. In FIG. 49, true structure (model 4 of 1LUD) is shown in black, initial structure (model 18 of 1BZF) is shown in gray, and optimum structure is shown in white. FIG. 50 shows 2, 4-diamino-5-(3,4,5-trimethoxy-benzyl)-pyrimidin-1-ium, which is a ligand of 1LUD.

Case 2)

1YER and 1YET, heat shock protein 90 (HSP90) have 100% homologous but are different conformations of active site because separate ligands bind to protein. Selecting

1YER which does not bind to ligand as initial structure, and using geldanamycin as a ligand, examination was made by comparison with 1YET which is true structure (FIG. 51). In FIG. 51, 3D structure of heat shock protein 90 is shown. Receptor (green A) and ligand (red B) of 1YET and receptor (blue C) of 1YER are shown.

In FIG. 52, result of normal mode analysis of 1YER is shown. In FIG. 52, intensity of fluctuation in dihedral angle ϕ (orange A), ψ (green B) is shown. The closer to 0.0 is intensity of the fluctuation, the stronger the constraint of dihedral angle becomes in the molecular dynamic (MD) calculation. Secondary structures determined by STRIDE are also shown: α-helix (red D) and β-sheet (blue D). The purple C shows an active site.

In FIG. 53 and FIG. 55, results of molecular dynamics with dihedral angle constrained using 1YER are shown. Rms with true structure, 1YET in the active site is shown. A is of main chain atom, B is of side chain atom, and C is of whole atom. FIG. 55 is a result of analysis of ligand binding to active site in 1YER. These are results of binding analyses when MD calculation was effected until 0.1 or 1.0 nanosecond, and dynamic structure of receptor are extracted at interval of 100 femtoseconds in the former and 100 and 1000 femtoseconds in the latter, and the sepalate populations are created by clustering those.

Evaluation was made based on rms with true structure of ligand in active site. In FIG. 54, parameter values for designating spatial points in ligand docking are shown. FIG. 54 shows information of structure-activity relationship obtained from 1YET.

In FIGs. 56 and 57, comparison between protein/ligand complex and true structure is shown. FIGs. 56 and 57 show binding of ligand in 1YER. FIG. 56 shows binding analysis between active site and ligand using a population created by clustering dynamic structure of receptor after those were extracted at interval of 100 femtoseconds within 0-1.0 nanosecond in the MD. Green A shows 1YET in true structure, blue B shows 1YER in initial structure, and red C shows optimum structure in ligand binding, "by element" color D shows ligand in true structure; and light blue E shows ligand by calculation result. Rms of ligand was 1.2081. By causing the ligand shown in FIG. 57 to bind, induction of 0.1619 by rms occurred in the main chain atom in the active site. FIG. 57 shows geldanamycin which is a ligand of 1YET.

Case 3)

1A9U and 1OUK, mitogen-activated protein kinase (MAP kinase), have 100% homologous but are different conformations of active site because separate ligands bind

to protein. Selecting 1A9U as initial structure, and using a ligand contained in 1OUK as a ligand, examination to compare with true structure of ligand in 1OUK was made (FIG. 58). In FIG. 58, 3D structure of mitogen-activated protein kinase is shown. Receptor (green A) and ligand (red B) of 1OUK and receptor (blue C) and ligand (light blue D) of 1A9U are shown.

In FIG. 59, normal mode calculation analysis of 1A9U is shown. In FIG. 59, intensity of fluctuation in dihedral angle ϕ (orange A), ψ (green B) is shown. The closer to 0.0 is intensity of the fluctuation, the stronger the constraint of dihedral angle becomes in the molecular dynamic (MD) calculation. Secondary structures determined by STRIDE are also shown: α-helix (red D) and β-sheet (blue D). The purple C shows an active site.

In FIG. 60 and FIG. 62, results of molecular dynamics with dihedral angle constrained using 1YER are shown. FIG. 60 shows a result of MD calculation of 1A9U. Rms with true structure, 1OUK in the active site is shown in FIG. 60. A is of main chain atom, B is of side chain atom, and C is of whole atom. FIG. 62 shows a result of analysis of ligand binding to active site in 1A9U. These are results of binding analyses when MD calculation was effected until 0.1 or 1.0 nanosecond, and dynamic structures of receptor were extracted at interval of 100 picoseconds in the

former and 100 and 1000 picoseconds in the latter, and three separate populations were created by clustering those. Evaluation was made based on rms with true structure of ligand in active site.

In FIG. 61, parameter values for designating spatial points in ligand docking are shown. FIG. 61 shows information of structure-activity relationship obtained from 1OUK.

FIGs. 63 to 65 show comparison between protein/ligand complex and true structure. FIGs. 63 to 65 show binding of ligand in 1AU9. FIG. 63 shows binding analysis between active site and ligand using a population created by clustering dynamic structures of receptor after those were extracted at interval of 100 femtoseconds within 0-1.0 nanosecond in the MD. Green A shows 1OUK in true structure, blue B shows 1A9U in initial structure, and red C shows optimum structure in ligand binding, "by element" color D shows ligand in true structure; and light blue E shows ligand by calculation result. Rms of ligand was 1.6112. By causing the ligand shown in FIG. 65 to bind, induction of 0.1871 by rms occurred in the main chain atom in the active site. In FIG. 64, true structure (1OUK) is shown in black, initial structure (1A9U) is shown in gray, and optimum structure is shown in white. FIG. 65 shows 4-[5-[2-(1-phenyl-ethylamino)-pyrimidin-4-yl]-1-methyl-4-(3-

trifluoromethylphenyl)-1H-imidazol-2-yl]-piperidine which is a ligand of 1OUK.

As shown in Cases 1) to 3), it was proved that a model of protein/ligand complex created by the ligand screening apparatus 100 can predict 3D structure of induced-fit type protein/ligand complex with high accuracy.

[Fourth example]

(Application example to in silico screening using Fxa)

By the ligand screening apparatus 100 in the above embodiment, a ligand having a possibility to bind to Fxa was screened from the compound database using 3D structure of Fxa (FIG. 66) which is one kind of serine protease. As 3D structure, 1AIX was used, and as a ligand database, 3633 kinds of ligands extracted from the PDB database was used. In accordance with the above embodiment, in silico screening was conducted. The results are shown in FIG. 67.

In FIG. 67, ligands to 100th from the top of interaction energy to 1AIX among the ligands in the compound database are shown. In FIG. 67, the bold character is a ligand contained in 1AIX, the italic character is serine protease. "PDB code" means code of PDB in which the ligand is originally contained. In FIG. 67, a ligand originally contained in 1AIX ranks 19th.

A protein/ligand complex structure in 19th ranking is

shown in FIG. 68 and FIG. 69. In FIG. 68, receptor is shown in white, and ligand of 1AIX is shown in black. FIG. 69 shows a ligand in 1AIX.

Every ligand in 35th, 38th and 80th rankings in FIG. 67 bind to serine protease.

These structures and protein/ligand complex structures are shown in FIG. 70 and FIG. 71, FIG. 72, FIG. 73, FIG. 74, and FIG. 75. In FIG. 70 and FIG. 71, structure of a protein/ligand complex in 35th ranking is shown. In FIG. 70, receptor is shown in white, and ligand of 1AUJ is shown in black. FIG. 71 shows a ligand in 1AUJ. FIG. 72 and FIG. 73, structures of a protein /ligand complex in 38th ranking is shown. In FIG. 72, receptor is shown in white, and true ligand (1FOR) is shown in black. The rms was 1.500. FIG. 73 shows a ligand in 1FOR. In FIG. 74 and FIG. 75, structures of a protein/ligand complex in 80th ranking are shown. As shown in FIG. 74, receptor is shown in white, and ligand of 1K1M is shown in black. Fig. 75 shows a ligand in 1K1M.

These results revealed that feasible compounds can be selected from the compound database according to the present invention.

[Fifth example]

(in silico screening under different conditions)

We verified that the ranking varies depending on the information of structure-activity relationship (SAR) by the ligand screening apparatus 100 in the above embodiment. We verified variation in ranking when the receptor is fixed.

Here, we executed in silico screening using protease of severe acute respiratory syndrome (SARS). As an initial structure, 1UK3 (B chain) not containing a ligand, and a binding mode of 1UK4 (B chain) including ligand was used as information of structure-activity relationship. The active site is a receptor residue region contained within 10 angstrom radius from each atom in the ligand of 1UK4 (B chain). As a ligand database, 3633 kinds of ligands collected from PDB were used. As a receptor dynamic structure cluster for use in binding analysis, the population assembling the structure extracted at interval of 100 femtoseconds within the range of 0-0.1 nanosecond was used. The calculation was conducted with the exclusion of hydrogen atoms.

FIG. 76 shows 3D structure of SARS protease. Receptor (green A) and ligand (red B) of IUK4 (B chain), as well as receptor (blue C) of 1UK3 (B chain) are shown.

In FIG. 77, a result of normal mode analysis of 1UK3 (B chain) is shown. Also intensity of fluctuation in dihedral angle ϕ (orange A), ψ (green B) is shown. The

closer to 0.0 is intensity of the fluctuation, the stronger the constraint of dihedral angle becomes in the molecular dynamic (MD) calculation. Secondary structures determined by STRIDE are also shown: α-helix (red D) and β-sheet (blue D). The purple C shows an active site.

In FIG. 78, a result of molecular dynamic calculation of 1UK3 is shown. Fig. 78 shows rms between MD result of 1UK3 (B chain) and 1UK4 (B chain) in the active site. A is of main chain atom, B is of side chain atom, and C is of whole atom.

Case 1) Designating four points in SAR

FIG. 79 shows spatial point designating within active sites of 1UK4. FIG. 79 is of structure-activity relationship information obtained from 1UK4(B chain). In FIG. 80, a result of in silico screening in 1UK3 (B chain) is shown.

In FIG. 81, 1UK3 are compared with 1UK4 which is a true structure. The ranking is 25th. Green A represents 1UK4 (B chain), bleu B represents 1UK3 (B chain) in initial structure, red C represents optimum structure in ligand binding, "by element" cooler D represents a true structure of peptide ligand (ASN-SER-THR-LEU-GLN) of 1UK4, and light blue E represents a calculated ligand. Rms of ligand was 2.5721. Rms with the true structure for main chain in active site was 1.0248 in initial structure, and

1.0792 in optimum structures.

In FIG. 82 and FIG. 83, the first ranking of in silico screening is shown. In FIG. 83, (C8-R) hydantocidin 5'-phosphate which is a ligand of 1QF4 is shown.

Case 2) Designating three spatial points in SAR

FIG. 84 shows spatial points designating within active sites of 1UK3. FIG. 84 shows structure-activity relationship information obtained from 1UK3 (B chain).

FIG. 85 shows a result of comparison between 1UK3 (B chain) and 1UK4 (B chain), and shows comparison of the optimum structure in 1UK3 with true structure. The ranking is 49th. Green A represents 1UK4 (B chain), bleu B represents 1UK3 (B chain) in initial structure, red C represents optimum structure in ligand binding, "by element" cooler D represents a true structure of peptide ligand (ASN-SER-THR-LEU-GLN) of 1UK4, and light blue E represents a calculated ligand. Rms of ligand was 2.0057. Rms with the true structure for main chain in active site was 1.0248 in initial structure, and 1.0469 in optimum structures.

In FIG. 86, a result of in silico screening executed while designating three spatial points of SAR is shown.

Case 3) Designating five spatial points in SAR

FIG. 87 shows spatial points designating within active sites of 1UK3. FIG. 87 is of structure-activity relationship information obtained from 1UK3 (B chain). In FIG. 88 is shown a result of in silico screening (for example, high throughput screening) executed for five designated spatial points of SAR.

FIG. 89 shows comparison between optimum structure and true structure in 1UK3. FIG. 89 shows a result of comparison between 1UK3 (B chain) and 1UK4 (B chain). The ranking is second. Green A represents 1UK4 (B chain), bleu B represents 1UK3 (B chain) in initial structure, red C represents optimum structure in ligand binding, "by element" cooler D represents a true structure of peptide ligand (ASN-SER-THR-LEU-GLN) of 1UK4, and light blue E represents a calculated ligand. Rms of ligand was 1.2578. Rms with the true structure for main chain in active site was 1.0248 in initial structure, and 1.1620 in optimum structures.

Case 4) Changing designated atom type of ligand atom

FIG. 90 shows spatial points designating within active sites of 1UK3. FIG. 90 is of structure-activity relationship information obtained from 1UK3 (B chain). In FIG. 91 is shown a result of in silico screening (for

example, high throughput screening) executed while changing the designated atom type of ligand atom.

FIG. 92 shows comparison between optimum structure and true structure in 1UK3. FIG. 92 shows a result of comparison between 1UK3 (B chain) and 1UK4 (B chain). The ranking is 774th. Green A represents 1UK4 (B chain), bleu B represents 1UK3 (B chain) in initial structure, red C represents optimum structure in ligand binding, "by element" cooler D represents a true structure of peptide ligand (ASN-SER-THR-LEU-GLN) of 1UK4, and light blue E represents a calculated ligand. Rms of ligand was 2.5216. Rms with the true structure for main chain in active site was 1.0248 in initial structure, and 1.0792 in optimum structures.

Case 5) Fixed receptor

FIG. 93 shows spatial points designated within active sites. FIG. 93 is of structure-activity relationship information obtained from 1UK4 (B chain). In FIG. 94 shows shown a result of in silico screening (for example, high throughput screening) executed while fixing the receptor.

FIG. 95 shows comparison between experimentally observed and calculated structures of ligand , when 1UK3 is superimposed on 1UK4. The ranking is 39th. Structure

of active site of 1UK3 is shown in gray, the former ligand is shown in black, and the latter ligand is shown in white.

As can be seen from Cases 1) to 4), the more SAR is designated, the better the ranking of the reference ligand is. That is, various ligands are caused to distribute in top ranking by conducting in silico screening with increased SAR information when binding information on reference ligand is reliable, and by reducing number of information of SAR and designating the relaxed range of atom types in ligand when the information is unreliable. More reliable result was produced when in silico screening was executed with the use of SRA information modified based on the distribution information.

Case 1) and Case 5) demonstrate variation in ranking depending on the presence/absence of dynamic structure of receptor. Optimization of structures of ligand and receptor is superior in preventing collision of atoms to optimization of structure of only ligand. Therefore, difference arose in optimization energy for placing ligand at the same position.

[Sixth example]

(Distribution of MD parameter concerning parameters of molecular dynamic calculation with dihedral angle contrained)

Now explanation will be given for distribution of parameter of MD with dihedral angle constrained in FMN-binding protein. Here, we verified whether similar results are obtained in molecular dynamic calculation with dihedral angle constrained and parameters of clustering even for NMR structure along with 1LUD by means of the ligand screening apparatus 100 of the aforementioned embodiment. We selected MODEL 1 of NMR structure (PDB code:1AXJ) of FMN-binding protein as an initial structure. As to the evaluation method, first example was followed except that parameters (α=80.0%, α=0.4 angstrom) for receptor dynamic structure clustering were fixed.

In FIG. 96, distribution of scores for determining parameter in molecular dynamic calculation with dihedral angle constrained in 1AXJ is shown. FIG. 96 shows distribution of parameter of MD with dihedral angle constrained in 1AXJ. The closer to A, the smaller the score is. As is the case with 1LUD, good results were obtained at maximum value 800 and minimum value 0 of dihedral angle constraint.

[Seventh example]

(MD with dihedral angle constrained)

Here, we verified dynamic structure of each atom by main chain in MD with dihedral angle constrained by means

of the ligand screening apparatus 100 in the aforementioned embodiment. There is sometimes the case that normal mode analysis does not converge and information of dihedral angle fluctuation is not obtained. Since good result is obtained when MD is conducted with a constant constraint (500) with respect to the main chain dihedral angle as shown in FIG. 13, we verified dynamic structure in this case. Following first example, MD was conducted without constraint, with constraint using dihedral angle fluctuation or with constant constraint (500). In FIG. 97 to FIG. 108, results of dynamic behavior in each atom in molecular dynamic calculation executed for 1LUD are shown.

FIG. 97 to FIG. 108 are results of MD calculation for 1LUD (MODEL 1). FIG. 97 and FIG. 98 are results of dynamic behavior of main chain atom in active site, FIG. 99 and FIG. 100 are of main chain atom in receptor, FIG. 101 and FIG. 102 are of side chain atom in active site, FIG. 103 and FIG. 104 are of side chain atom in receptor, FIG. 105 and FIG. 106 are of whole atom in active site, and FIG. 107 and FIG. 108 are of whole atom in receptor. For each model structure of 24 kinds described in PDB file of 1LUD, rms with main chain atom in active site by MD of MODEL 1 was calculated, and an average rms was displayed by dotted lines. Difference of main atom in the active

site from its initial structure was shown by rms in the presence (A) or absence (B) of dihedral angle constraint (A), or with constant dihedral angle constraint (C).

Although not shown herein, with regard to 1CBQ, 1J9G, 1MMB, 1BZF (MODEL 18), 1YER, 1A9U and 1UK3 (B chain), the results of constrained MD based on fluctuation of main chain dihedral angle demonstrated that a side chain atom not being constrained also exhibited a certain measurement of motion even if that of the main chain atom was constrained as is the case with FIGs. 109 to 111. It can be understood that motion of main chain atom heavily influences on the motion of the receptor.

[Eighth example]

(Binding analysis under different conditions)

Here, we verified that induction was caused even when different parameters were used in MD with dihedral angle constrained and clustering by means of the ligand screening apparatus 100 in the aforementioned embodiment. Second example was followed except that the maximum value of constraint was set at 100, the minimum value was set at 0, the clustering coefficients α and β of receptor dynamic structure were set at 80.0% and 1.0 angstroms, respectively. For clustering of receptor dynamic structure, the structures were extracted at a interval of

100 femtoseconds within the range of 0 to 0.1 nanosecond was used, and a polulation of those was created. Receptor residues within 6 angstroms radius from each atom in the ligand are defined to form an active site.

In FIG. 109 to FIG. 111, results of receptor/ligand binding under different conditions are shown.

(i) In 1BZF (MODEL 18), binding of ligand caused induction of 0.2686 in rms of main chain atom in the active site (FIG. 109). In overall rms of active site, induction of 0.1224 was caused. Rms of ligand was 0.8526.

(ii) In 1YER, binding of ligand caused induction of 0.22376 in rms of main chain atom in the active site (FIG. 110). In overall rms of active site, induction of 0.0816 was caused. Rms of ligand was 0.7246.

(iii) In 1A9U, binding of ligand caused induction of 0.2150 in rms of main chain atom in the active site (FIG. 111). In overall rms of active site, induction of 0.0464 was caused. Rms of ligand was 0.9464.

Green lines show true structure, blue lines show initial structure, red lines show optimum structure, "by element" color lines show true ligand, and blue lines show optimum ligand.

As shown in FIG. 109 to FIG. 111, optimum results were obtained within a given condition even if each condition differs.

[Ninth example]

(Binding analysis when true structure is selected as initial structure)

Here, since binding modes of 1BZF and 1LUD of DHFR resemble, binding analysis of ligand of 1BZF was conducted with partial modification of structure-activity relationship information by means of the ligand screening apparatus 100 in the aforementioned embodiment. Condition was such that 1BZF (MODEL 18) was used as initial structure, a cluster created from a population of 0 to 0.1 nanosecond was used. In FIG. 112, spatial point designated in active site of 1BZF is shown. FIG. 112 is a view of structure-activity relationship information modified for 1BZF. In FIG. 113 and FIG. 114, results of receptor/ligand binding in 1BZF are shown. In FIG. 113 and FIG. 114, results of ligand binding analysis of 1BZF (MODEL 18) are shown.

(i) As optimum structure of a receptor , initial structure was selected. Rms of ligand was 0.8884 (FIG. 113).

(ii) Trimetrexate, which is a ligand for 1BZF(MODEL 18) (FIG. 114).

Since the initial structure was a structure that was originally registered in PDB, namely optimum structure, the calculation results could be reproduced as shown in

FIG. 113 and FIG. 114.

INDUSTRIAL APPLICABILITY

As described above, a ligand screening apparatus, a ligand screening method, and a program and recording medium according to the present invention seem to be very useful in the fields conducting analysis of receptor/ligand binding (drug design), especially for molecular design of pharmaceutical and agricultural chemicals. The present invention can be widely practiced in many industrial fields, especially in pharmaceutical, food, cosmetics, medical, structural analysis, functional analysis and the like fields, and hence are very useful.